INVITED SPEAKERS

Prof. Sisir Roy

National Institute of Advanced studies, Indian Stastical Institute, Kolkata.

Sisir Roy, the Visiting Chair Professor at NIAS (National Institute of Advanced Studies). He has been a professor at Indian Statistical Institute, Kolkata for 21 years and also a visiting professor at University of Arkansa, George Mason University and visiting scientist at Henri Poincare Institute. His areas of research include Foundations of Quantum Mechanics, Brain Function modeling and Cognitive Science, Data Analysis and Quasar Astronomy.

Prof. John Bryden

PM University, Saudi Arabia.

John Bryden is an algebraic topologist whose primary specialty is homotopy theory, although he is also interested in algebraic geometry, low dimensional topology, geometric topology, representation theory and the new subjects of quantum topology and link homology. The fundamental theme of his research has been the application of algebraic topology to low dimensional manifolds.

Prof. M. Ramachandran

Leeds Beckett University, UK.

Muthu Ramchandran is currently a Principal lecturer in Computing and creative Technologies School as part of the faculty of arts. Environment and Technology at Leeds Beckett University, UK .Previously, he spent nearly eight years in industrial research (Philips Research Labs and Volantis system Ltd. Surrey UK) where he worked on software on software architecture, reuse, and testing.

Dr. Om D Deshmukh

Senior Research Scientist, Area Manager Multimedia Analytics, Xerox Research Center India, Bengaluru.

Om Deshmukh is a senior Research Scientist at XRCI and part of the Analytics and Big Data team. He is particularly interested in technological research that leads to high business and societal impact in the emerging markets. Om Deshmukh's PhD student at IISc (co-advised with Dr. PrasantGhosh) receives the Prime Minister's Fellowship for Doctoral Research. XRCI Education team, working in close collaboration with the Education team at PARC-E, won the Mass EduData Challenge in two categories.

Dr. T.R. Gopalakrishnan Nair

Rector, RRGI, Bengaluru.

A visionary technologist, contributed extensively in Education, Advanced Research and Administration throughout his career. He is a senior member of IEEE for last two decades and a senior member of various other societies like ACM (USA) and CSI India. The latest is the recognition from Elsevier, Amsterdam recognizing him with the award of "Outstanding Contributions in Reviews", which puts him in the top 10% of the experts. He was given the Glory of India Award by IIFS in July 2013 in Bangkok.

Mr. Guru Ragaventhiran. R

Practice Manager, Management Consulting, KPMG, Bengaluru.

Guru is a Technology leader with 14 years of experience in Oracle Applications consulting including people soft, Oracle cloud applications. He has an excellent blend of consulting, Project management and leadership skills with extensive experience in pre-sales, solutioning, delivery of projects, Team building and Development activities. He has international success working on-site with customers in assisting them in their business modernization journey.

Dr. Pethuru Raj Chelliah

Infrastructure Architect, IBM Global Cloud Center of Excellence, IBM India, Bengaluru.

Pethuru Raj has been working as a cloud infrastructure architect in the IBM Global Cloud Center of Excellence (CoE), IBM India Bangalore. Previously he worked as TOGAF-certified enterprise architecture (EA) consultant in Wipro Consulting Services (WCS) Division, Bangalore. He also had a fruitful stint as a lead architect in the corporate research (CR) division of Robert Bosch, India. He has gained more than 16 years of IT industry experience.

Dr. Manjaiah

Professor & Chairman of BoS both UG & PG in Computer Science, Mangalore university, Mangalore.

Presently he is working as a Professor & Chairman of BoS both UG & PG, Department of Computer Science, Mangalore university, Mangalore. He is Member of Board of Studies / Board of Examiners and Doctoral Committee in Computer Science / Information Science & Technology / Bioinformatics of other reputed Universities and Institutions in India. He is also an Advisory and Expert committee member of AICTE, UGC, Govt. of India and KPSC, Govt. of Karnataka. He has received BEST CITIZEN of INDIA for the year 2013 by International publishing house, NEW DELHI.

RAJARAJESWARI COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Accredited by NBA)

1st International Conference on

Innovations in Computing and Networking (ICICN-16)

on 12th & 13th May 2016

11th May 2016 Pre-Conference Tutorials

Venue:

RajaRajeswari College of Engineering

Ramohalli Cross, Kumbalgodu, Mysore Road, Bengaluru-560074, Karnataka.

INAUGURAL FUNCTION

Cordially invites you to grace the inaugural function

ICICN-16

The Keynote Address

"Quantum Probability, Bayesian Brain and Computation"

Delivered by

Prof. Sisir Roy

Chair Professor, National Institute of Advanced Studies, Bengaluru.

Former Professor, Indian Statistical Institute, Kolkata.

Presided by

Sri A.C. Shanmugam, Chairman, RRGI.

Chief Patron

Sri A.C. S. Arun Kumar, Vice-Chairman, RRGI.

Patron

Sri. S. Vijayanand, Executive Director, RRGI.

General Chair

Sri. S. Jeyabalan, Special Officer, RRGI.

Dr. R. Balakrishna, Principal, RRCE.

Program Chair

Dr. T.R. Gopalakrishnan Nair, Rector, RRGI.

Organizing Secretary

Dr. S. Usha, Prof. & Head, CSE, RRCE.

VALEDICTORY FUNCTION

Valedictory Address by

Dr. T.R. Gopalakrishnan Nair

Rector, RRGI, Bengaluru.

PANEL DISCUSSION

12th May 2016

3:00 p.m - 4:00 p.m @ A.P.J. Abdul Kalam Hall

Topic: "IOT - Challenges and Proliferation"

Moderator: Dr. T.R. Gopalakrishnan Nair

Rector, RRGI, Bengaluru.

Panelists

Prof. John Bryden PM University, Saudi Arabia.	Quantum Computing
Prof. M Ramachandran Leeds Beckett University, UK.	Cloud Computing
Dr. Pethuru Raj Infrastructure Architect, IBM Global Cloud Center of Excellence,IBM India, Bengaluru.	Internet of Things
Dr. Sathyan Head, Department of Physics, RRCE, Bengaluru.	Physics of Computing

Tutorial Sessions

11th May 2016

10.15 a.m. - 11.15 a.m. Dr. A.P.J. Abdul Kalam Conference Hall	"Software Engineering for Service and Cloud Computing" **Prof. M. Ramachandran** Leeds Beckett University, UK.
11.30 a.m. - 12.30 p.m. Dr. A.P.J. Abdul Kalam Conference Hall	"Introduction to Topological Quantum Field Theories" **Prof. John Bryden** PM University, Saudi Arabia.
12.30 p.m. - 1.30 p.m. Dr. A.P.J. Abdul Kalam Conference Hall	"Multimedia Analytics," **Dr. Om D Deshmukh** Senior Research Scientist, Area Manager Multimedia Analytics, Xerox Research Center, Bengaluru.
2.30 p.m. - 3.30 p.m. Dr. A.P.J. Abdul Kalam Conference Hall	"Cloud and Its Applications" **Mr. Guru Ragaventhiran.R** Practice Manager, Management Consulting, KPMG, Bengaluru.
3.30 p.m. - 4.30 p.m. Dr. A.P.J. Abdul Kalam Conference Hall	"Enterprise Architecture for Large Scale Reuse" **Prof. M. Ramachandran** Leeds Beckett University, UK.

RAJARAJESWARI COLLEGE OF ENGINEERING

NBA Accredited & ISO 9001 : 2008 Certified Institution
No.14, Ramohalli Cross, Kumbalgodu, Mysore Road
Bangalore-560 074

Department of Computer Science and Engineering

1st International Conference on Innovations in Computing and Networking (ICICN - 16)

12-05-16 & 13-05-16

International Journal Of Advanced Networking Applications (IJANA)

Conference Schedule

12th May 2016

Sl.No	Time	Event	Venue
1	08:30AM - 09:30AM	**REGISTRATION**	RRCE, Ground Floor ,Lobby
2	09:30AM - 10:00AM	**INAUGURATION** **Prof. Sisir Roy** Chair Professor, National Institute of Advanced studies, Former Professor, Indian Statistical Institute, Kolkatta **Prof. Muthu Ramachandran** Leeds Metropolitan University, UK **Prof. John Bryden** PM University, Saudi Arabia	Dr. A. P. J. Abdul Kalam Hall
3	10.00 - 10.30AM	**Keynote Address** by **Prof. Sisir Roy** on "Quantum Probability Bayesian Brain"	
4	10.30-11.00AM	**Special Address** on "Computing Technologies, Nuromorphic & Quantum Visions" by **Dr. T.R. Gopalakrishnan Nair** Rector, RRGI, Bengaluru	

5	**11.00-11.15AM**	**Tea Break**	
6	**11.15AM-12.45PM**	**SESSION 1**	**CSE Seminar Hall** **ECE Seminar Hall** **Dr. A. P. J. Abdul Kalam Hall** **Smart Class room no.522**
7	**12.30AM-1.30PM**	**Invited Talk** **by** **Prof. Pethuru Raj** **Cloud infrastructure architect, IBM India, Bengaluru** **on** **"Envisioning Real-time IoT Data Analytics through Fog / Edge Device Clouds"**	
8	**1.30-2.30PM**	**LUNCH BREAK**	**Room No. 504**
9	**2.30-4.00 PM**	**SESSION 2**	**CSE Seminar Hall** **ECE Seminar Hall** **Dr. A. P. J. Abdul Kalam Hall** **Smart Class room no.522**
10	**4.00-4.45PM**	**Special Talk** **by** **Prof. Manjaiah** **Professor & Chairman, Dept. of CSE, Mangalore university, Mangalore** **on** **"Importance of Open Source Software [oss] Computational tools for Researchers"**	**Dr. A. P. J. Abdul Kalam Hall**
11	**4.45-5.30PM**	**Panel Discussion Followed by High Tea**	**Dr. A. P. J. Abdul Kalam Hall**

Panel Discussion

Topic: "IoT –Challenges and Proliferation"

Moderator: Dr. T.R. Gopalakrishnan Nair
Rector, RRGI, Bengaluru

PANELISTS

Prof. John Bryden PM University, Saudi Arabia	**Quantum Computing**
Prof. Muthu Ramachandran Leeds Metropolitan University, UK	**Cloud Computing**
Prof. Pethuru Raj Cloud infrastructure architect, IBM India, Bengaluru	**Internet of Things**
Prof. Satyan Head, Dept. of Physics, RRCE, Bengaluru	**Physics of computing**

RAJARAJESWARI COLLEGE OF ENGINEERING

NBA Accredited & ISO 9001 : 2008 Certified Institution
No.14, Ramohalli Cross, Kumbalgodu, Mysore Road
Bangalore-560 074

Department of Computer Science and Engineering

1st International Conference on Innovations in Computing and Networking (ICICN - 16)

12-05-16 & 13-05-16

International Journal Of Advanced Networking Applications (IJANA)

Conference Schedule

13th May 2016

Sl.No	Time	Event	Venue
1	9.30-10.30AM	Special Talk by Prof. John Bryden PM University, Saudi Arabia On "An Introduction to Mathematics of Quantum Computing"	Dr. A. P. J. Abdul Kalam Hall
2	10.30-10.45AM	Tea Break	
3	10.45-12.15PM	Session 3	CSE Seminar Hall ECE Seminar Hall Dr. A. P. J. Abdul Kalam Hall Smart Class room no.522
4	12.20-1.00PM	Special Talk By Prof. John Bryden PM University, Saudi Arabia On "Braid representation theory"	Dr. A. P. J. Abdul Kalam Hall
5	1.00-1.45PM	Lunch Break	Room No. 504
6	1.45-3.15PM	Session 4	CSE Seminar Hall ECE Seminar Hall Dr. A. P. J. Abdul Kalam Hall Smart Class room no.522
7	3.15-4.00PM	VALEDICTORY Dr. T.R. Gopalakrishnan Nair Rector, RRGI, Bengaluru. Dr. R Balakrishna Principal, RRCE	Dr. A. P. J. Abdul Kalam Hall

Programme Schedule | Day – 1 | Theme: NETWORKING & CLOUD COMPUTING Venue: CSE Seminar Hall

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 1:- 11.15 am – 12.45 pm	ICICN16-CS004	Improving the Wireless Sensor Network Lifetime by Diminishing the Energy Consumption via Nature Inspired Technique	Vinay R N Parvathi C	**Dr. Manjaiah** Mangalore University	**Dr. Cyril Raj** Dr. MGR University, Chennai
		ICICN16-CS007	An Analytical Framework Intended For Congestion in Wireless Networks	Daina K K, Dhanalakshmi H.B , Pallavi Gowdoor , Sindhura D		
		ICICN16-CS012	Three Tier Approach for Secure Data Transmission by Using Steganography, Logistic Maps with Genetic Algorithm and Visual Cryptography	Vamsi Krishna.Y, Varuni.R.V , Shivaranjini.A		
		ICICN16-CS015	Greedy Forwarding for Wireless Sensor Networks with Guaranteed Delivery	Namita Mirjankar Siddhitha S Pai , Lohit Patil		
		ICICN16-CS020	A Survey On SOA Governance Scope, Objectives And Its Policies	Pushpalatha B Prof R Jaya		
		ICICN16-CS025	An Efficient Reduction of Encryption Keys for Group Data Sharing via Cloud Storage	ManjulaK.,MahalakshmiG., YashaswiniM.S.		
		ICICN16-CS031	Provision of Securing Data in Cloud through Auditing and Deduplication	Dhananjaya V Naveen Ghorpade		
		ICICN16-CS037	A Third Generation Design For Secured Atm Transaction Using Sim Card	A.Chandhanaa Durga.S		
		ICICN16-CS046	Cloud Latency: A Cloud Based Latency Checking System for the Developers	Arun Kumar P L Mr.Arun Kumar D R		
		ICICN16-CS050	Traffic Profiling and Dynamic weight computing using dWRR for Average packet delay versus traffic load	AkshathaKamath HumeraSiddiqua Usha M. S		

Programme Schedule | Day – 1 | Theme: DATAMINING & BIGDATA **Venue:CSE Smart Room (No. 522)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 1:- 11.15 am – 12.45 pm	ICICN16-CS003	A Survey On Dynamic Web Services Composition Existing Methods Related Techniques	NehaSinghal Dr.Y.S KUMARASWAMI	Prof. Rajesh K S RRCE, Bengaluru	Dr. D. V. Ashok JSS, Bengaluru
		ICICN16-CS040	An Approach to Mining Sequential Rules Common to Several Sequences	G.Jayagopi Dr.S.Pushpa		
		ICICN16-CS063	Weighted Item Set Mining from Bigdata using Hadoop	Divya.M.G , Nandini.K Priyanka.K.T ,Vandana.B		
		ICICN16-CS072	Live Website Traffic Analysis Integrated with Improved Performance for Small Files using Hadoop	Navyashree B L, Rakshitha H , Ranjitha M, T Auntin Jose		
		ICICN16-CS082	Customer behavior analysis of web server logs using Hive in Hadoop Framework	Lavanya KS Srinivasa R		
		ICICN16-CS083	Ranking Detection And Avoidance Frauds Mobile Apps Store	Abhiilash T P L Dinesha		
		ICICN16-CS085	Stream Processing of Scientific Big Data on Heterogeneous Platforms with Image Analytics on Big Data	Ms. Shruti B Karki Mr. Auntin Jose. T		
		ICICN16-CS094	Analysis of DWT OFDM using Rician Channel and Comparison with ANN based OFDM	Geeta S H Smitha B		
		ICICN16-CS132	Automatic range timeline generation for volume based tweet stream	Tejaswini K, Madhavi H Sreenivasa B R		
		ICICN16-CS136	Overview of Big Data Visualization	SagargoudaS.Patil		
		ICICN16-CS202	Anti-Jamming Technology For Mobile Signals Using Emp Jammer	Akash R Mannari		

Programme Schedule | Day – 1 | Theme: IMAGE PROCESSING **Venue: ECE Seminar Hall (4th Floor)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 1:- 11.15 am – 12.45 pm	ICICN16-CS006	Analysis of White Blood Cells for Malaria Detection	Hemalatha K L, Karthik R, NandanSateeshHegde	**Prof. Bhuvaneshwari** RRCE, Bengaluru	**DR. J. Prakash** BIT, Bengaluru
		ICICN16-CS008	3D Reconstruction Methodologies: a Review	Sowmya S G Dr. Sunanda Dixit		
		ICICN16-CS018	Qualitative Improvement in Wireless Capsule Endoscopy Images using Specific Contrast Enhancement Techniques	Muddasir Khan , Ashwini S.S M Z Kurian		
		ICICN16-CS102	Intelligent Software Agent based Image Fusion in Wireless Multimedia Sensor Network	Janaki K, Dr.M.Madheswaran		
		ICICN16-CS026	State of Art Technique that Accomplishes Digital Image Protection and Self Recovery	Prema C.R Megha S		
		ICICN16-CS043	Super Resolution Reconstruction Based On Different Techniques Of Registration And Interpolation	KamalavvaTotagiChidananda Murthy M V M Z Kurian		
		ICICN16-CS053	A Survey on Different Methods for Brain Tumor Segmentation	NgangomPriyobata Singh Dr. Sunanda Dixit Prof. B.I. Khodanpur		
		ICICN16-CS206	Emotion Recognition: Detecting Emotions From Textual Documents, Blogs And Audio Files	Lakshmi.K ,Harshitha.K.Rao Revathi.V, Mrinal, Archana.T.P		
		ICICN16-CS069	Speech Controlled Smart Wheelchair Using Regional Languages	Chandramma R , Akshatha K Gowda Asha U ,Jyothi B S Meghana B Bapat		
		ICICN16-CS076	Study and Analysis of DWT-SVD Based Digital Image Watermarking Technique for colour images	Rebecca A Dr. UshaSakthivel		

Programme Schedule | Day – 1 | Theme: AI & IoT **Venue: ECE Seminar Hall (4th Floor)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 2:- 2.30 pm – 4.00 pm	ICICN16-CS014	A Survey On Importance And Applications Of Data Mining Using Artificial Neural Network	Ms. R.Jaya Mr.Giridhar V	Dr. P.V. Rao RRCE, Bengaluru	Dr. Mohan H.S. SJBIT, Benaluru
		ICICN16-CS017	A Survey Of Various Technologies For Soa Adopted By Software Industries	Karthik .V Prof. R.Jaya		
		ICICN16-CS019	Prominence of Expert System and Case Study- DENDRAL	NamitaMirjankar ShrutiGhatnatti		
		ICICN16-CS024	Study of IoT: Understanding IoT Architecture, Applications, Issues and Challenges	Soumyalatha Shruti G Hegde		
		ICICN16-CS045	Methodology To Make Natural Language as Computer Programming Language	Nalina H J Mr.Vinay T R		
		ICICN16-CS092	A Novel approach to design a Smart bin through IoT	Sruthi K V Manjunath K N		
		ICICN16-CS107	Preparation of Papers for An E-Governance Portal for Tax Payment	Rizwanaunnisa Bhanu K. N.		
		ICICN16-CS109	Social Media Analytics for Business Analysis and Improved Decision Making	HarshaVasishta M. Sreenivasa B. R.		
		ICICN16-CS114	Internet of Things for Environmental Monitoring	Srinivasa.R, Bhoomika.K.N Deepa.C, Rashmi.R.K		
		ICICN16-CS127	A Novel approach of 4*4 Vedic Multiplier using reversible logic gates	shivarathnamma G.Jyothi Dr. M Z Kurian		
		ICICN16-CS225	Enhanced AODV with Secrete Key Sharing to Improve Security and Energy Efficiency in MANETs	Murali G , Divyashree B.N , Dr. Balakrishna R Dr M.Vinayaka Murthy		

Programme Schedule | Day – 1 | Theme: NETWORKING & CLOUD COMPUTING Venue: CSE Seminar Hall

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 2:- 2.30 pm – 4.00 pm	ICICN16-CS056	A Cloud-Based Venue Recommendation Framework on Mobi Context	Ambika L G Anitha T N	**Dr. Malathy** RRCE, Bengaluru	**Dr. Shadaksharappa** SEC, Anekal, Bengaluru
		ICICN16-CS057	A Novel Identity-Based Encryption using Outsourced Revocation in Cloud Computing	Meenakshi.S ,Radhika.B.R Sripriya.B.S Tejashwini Nandini.G		
		ICICN16-CS060	Portable and Inevitable Path through Versatile Information Gathering in Remote Sensor Systems	Rekha B SowmyaNaik P.T		
		ICICN16-CS061	Advanced Network Monitoring System Using A Hybrid Of Agent And Agentless Concepts	Rahul S Sreedhar, PrithviVihari R Dr. N Guruprasad		
		ICICN16-CS067	Securing The Data Confidentiality of Patients in Distributed M-Healthcare Cloud Computing System	Yamuna R , Chandrika Mahima J, Swetha P		
		ICICN16-CS081	Three Party Authentication Using Quantum Key Distribution protocol	Aneesa C V , Drishya Nair NithaKamar , Pratap M S Ponnu Narayanan		
		ICICN16-CS086	Reducing the Time Factor for analyzing the Errors and Fixing in Big Sensor data	Anusha G Dr. N Guruprasad		
		ICICN16-CS087	Automating Infrastructure as a Code using Continuous Integration and Continuous Delivery	Bhanupriya H, Ravi ShekharJha ,Dr.Krishna A		
		ICICN16-CS097	Privacy Preserving Mobile Access of Health Data Enabling Auditability	Deepashree. N ,Divyashree Tanushree K N , Vinutha H		
		ICICN16-CS112	EAMC: using Fog Computing and RRP for Exigency Alert Service	Srinivasa R Deepika M		

Programme Schedule | Day – 1 | Theme: IMAGE PROCESSING **Venue: CSE Smart Room (No. 522)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
12th MAY 2016	Session – 2:- 2.30 pm – 4.00 pm	ICICN16-CS089	Enhancing Image Copy-Move Forgery Detection using Particle Swarm Optimization Techniques	Manjesh BN Madhu BN	**Prof. Bhanu K N** RRCE, Bengaluru	**Dr. G. T. Raju** RNSIT, Bengaluru
		ICICN16-CS091	A Survey on Different Methods for Liver Segmentation	MallikarjunKesaratti Dr. Sunanda Dixit Prof. B.I. Khodanpur		
		ICICN16-CS095	Certain investigations on filter performance for skin texture analysis	Dr.Punal M Arabi , Gayatri Joshi , KavyaParameswaran Abhik Raj Subedi		
		ICICN16-CS096	Approaches To Content-Based Image Retrieval	Lavanya B , Meghashree M D Sanjana P ,Usha M Poonamkumari		
		ICICN16-CS104	A Survey on usage of FMRI technique in BCI	Gireesh H R Dr.B.S.Pradeep		
		ICICN16-CS108	Multi Model Image Registration and Fusion using Fast Discrete Contourlet Transform	Yashaswini V A P. Bhuvaneswari ,B.Aravind		
		ICICN16-CS110	An Effective & Automated MR Brain Image Segmentation & Tumor Detection	PushpaLatha N , Aruna Kumara B , Shashidhar V, Bharath J		
		ICICN16-CS124	Classification Of Kidney Stone Using Glcm	DrPunal.M.Arabi, Surekha Nigudgi , Rohith.N.Reddy Dhatri		
		ICICN16-CS139	Secured Secret Image Transmission by using Fragment Visible Mosaic Image Technique	Ashalatha H M , M.D Anitha Devi , Dr. M. Z Kurian , Dr. K.B Shivakumar		
		ICICN16-CS141	A Framework for Text Analytics using the Bag of Words (BoW)Model for Prediction	DeepuGowda , Pethuru Raj S.Rajaraajeswari		

Programme Schedule | Day – 2 | Theme: NETWORKING & CLOUD COMPUTING Venue: CSE Seminar Hall

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 3:- 10.45 am – 12.15pm	ICICN16-CS113	Technique to balance energy efficient clustering with data transmission in large scale sensor network	Jyothi A.P Dr. UshaSaktivel	**Dr. Arunachalam** RRCE, Bengaluru	**Dr. H.S. Guruprasad** BMSCE, Bengaluru
		ICICN16-CS115	Efficient Energy Conservation Techique Using Sink Node Mobility For Big Data Gathering In Densely Distributed Wsn	Mrs.Anitha K Karthik.K.J		
		ICICN16-CS121	Secured Load Re-Deployment of File Chunks in Distributed File Systems	Roopadevi D.S Mrs. Nandini G		
		ICICN16-CS149	Analysis And Implementation Of Multicast Video Using Optical Network By Streaming Process.	Nalinakumari Bhargav Rama Gowd Hitesh Mehta		
		ICICN16-CS151	Clustering In Mobile Wireless Sensor Networks: A Review	Shantala Devi Patil Vijaya Kumar B P		
		ICICN16-CS152	FIELD weakening Strategy in Vector Control of Induction Motor during Inverter Voltage Saturation	Pallavi H.R. Sumitha T. L		
		ICICN16-CS162	Investigating Hackers On Facebook Application Using FRAppE	Asha Alias ,Rincy Varghese Ritu.M.Varghese V.M.Saravanaperumal		
		ICICN16-CS165	A Novel Approach for Efficient Data Collection in Cluster Based Communication using Compressive Sensing and MIMO Techniques to Extend Network Lifetime	SunandaPatne Vinodha.K		
		ICICN16-CS176	Management of QoS in Sensor Cloud	Anasuya N J, Akhila SJ		
		ICICN16-CS178	Controlled Data Access Using IBE in Cloud Computing	Bhoomika M U , Chandini Harshitha.R.S ,Suraj S Gurav Prof. Mangala.C.N		

Programme Schedule | Day – 2 | Theme: DATAMINING & BIGDATA **Venue: CSE Smart Room (No. 522)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 3:- 10.45 am – 12.15pm	ICICN16-CS138	Predictive, Corrective actions of Energy management and benefits through Data Analytics with Intelligent Reporting	Prof,Shamshekhar. S. Patil, Spurthi. R .Terwai	**Dr. Senthil Kumarn** ACSCE, Bengaluru	**Dr. Sheela T** SSEC, Chennai
		ICICN16-CS143	Implementation of an Efficient MongoDBNoSQLExplorer for Big Data Visualization	Chaitanya.P ,Ranjan H.P Kiran T.S, Anitha.K		
		ICICN16-CS156	Evaluating the Job Performance using DyScale Scheduler and MapReduce in Hadoop framework	Supriya.R Kantharaju.H.C		
		ICICN16-CS161	Implementation of Extended MapReduce for Emerging BigData	Nayana N kumar Jayashree L K		
		ICICN16-CS163	A Friend Recommendation System based on Similarity Metric and Social Graphs	Rashmi. J Dr. Asha. T		
		ICICN16-CS170	A Selective Approach for Storing Small Files in Respective Blocks of Hadoop	Chethan.R , ChandanKumar JayanthKumar.S, Girish.H.J , Prof. Mangala.C.N		
		ICICN16-CS189	Integration of Visual Temporal and Textual Distribution Information for News Video Mining	Prof. Shivamurthy RC Tauseef Ahmed S S		
		ICICN16-CS010	Automatic road extraction from satellite image by Difference of Gaussian and convolution overlap add method	Veda.D, Suma B Rao		
		ICICN16-CS177	Steganographic Data Embedding with Revocable Texture Synthesis	AkshayU.Joshi Sachin K S ,Vishal .M Aruna Kumara B		
		ICICN16-CS021	Review Paper On Photonic Crystal Basedoptical Memory	Jyoti. B		

Programme Schedule | Day – 2 | Theme: IMAGE PROCESSING **Venue: ECE Seminar Hall (4th Floor)**

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 3:- 10.45 am – 12.15pm	ICICN16-CS145	A New Approach of Image Steganography Technique for Information Hiding using NearestFilling Technique	S. Sreedharkumar Bhagyalakshmi	**Dr. C. S. Pillai** ACSCE, Bengaluru	**Dr. Swarna Jyothi** RRCE, Bengaluru
		ICICN16-CS155	Splendid Resolution Images Using Contrast Limited Adaptive Histogram Equalization	Rajesh K.S Sushmitha S.R		
		ICICN16CS173	Identifying suitable enhancement technique for thermal and non thermal diabetic foot images	Dr.Punal M Arabi Gayatri Joshi ,TejaswiBhat Harsha L Singh		
		ICICN16-CS174	A Framework for Anomaly Event Detection by Analysing the Video Sequences	Puttegowda D U N Sinha		
		ICICN16-CS180	Gesture Recognition Based Car Gaming	RubeenaMuheeb Dr. Parameshachari B D Nagashree R N, Deekshith B N, Keerthikumar M, Rashmi		
		ICICN16-CS181	ECG Signal Steganography Using Wavelet transforms	Asha N S, Anithadevi M D Dr. Shivakumar K B Dr. M Z Kurian		
		ICICN16-CS186	Manual Segmentation Of Hrct Lung Image Using Gap Filling Technique	DrPunal.M.Arabi VamshaDeepa.N Lakshmi Bai P		
		ICICN16-CS198	Automated Attendance System Using Image Processing	Pooja G.R ,Poornima M Palakshi S, M Bhanu prakash Verma		
		ICICN16-CS062	An efficient Image Denoising using Epitome and Weiner filter	Ms.Janaki Ranjith R D		
		ICICN16-CS135	Authenticating Login session using Mouse biometric with AES Encryption	Shruthi G Shilpa S.G GeetaSHukkerimath		

Programme Schedule | Day – 2 | Theme: AI & IoT **Venue: ECE Seminar Hall (4th Floor)**

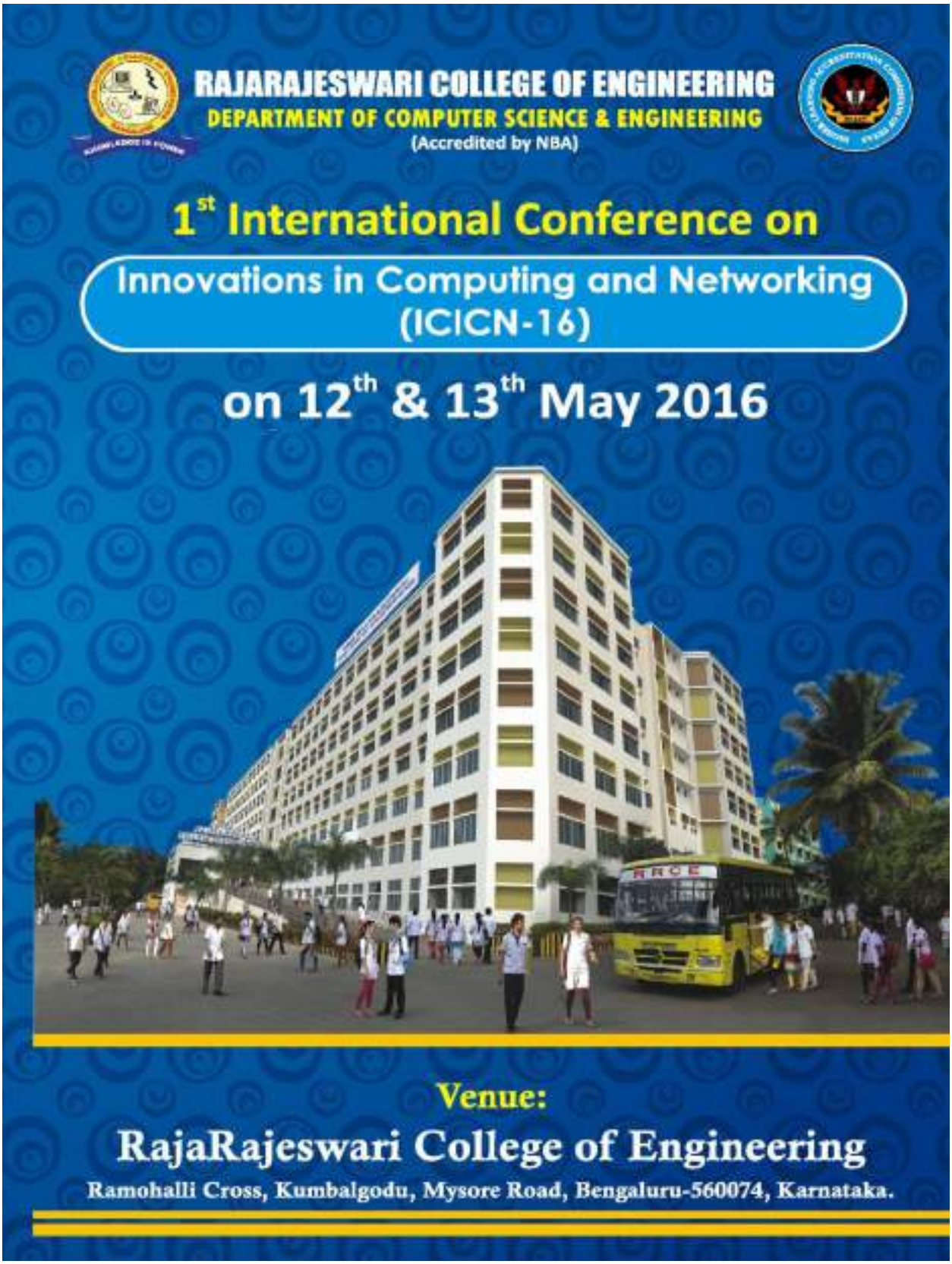
Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 4:- 1.45 pm – 3.15pm	ICICN16-CS129	Design, Implementation and comparison of 8-Bit Vedic Multiplier Using Multiplexers and Logic Gates	Sumanashree C.V M.C.Chandrashekhar M Z Kurian	**Dr. Rangaiah** RRCE, Bengaluru	**Dr. Meenakshi Sundaram** DBIT, Bengaluru
		ICICN16-CS150	An Generic Cloud Framework for Cloud Based Applications	Arvind Kumar M MrT.Auntin Jose		
		ICICN16-CS158	An Interactive Smart Glass	Srinivas M , Gangaram D Varun Reddy R, AtulKewat K Dr. UshaSakthivel		
		ICICN16-CS171	Design Approach for Increased Lifetime of WSN using Artificial Neural NYetwork Based Data Aggregation	Manomi K S , Vinutha C B M Z Kurian		
		ICICN16-CS160	A Novel Frame Work for E-Ndorse	Divya R K Bhanu K N		
		ICICN16-CS009	Audio-video Geners Classification Using AANN	K. Subashini		
		ICICN16-CS100	Approaching Machine Learning Using H2O and Inspecting it on Apparel Industry Using D3 Visualization	Chaitra D.B Mrs.Bindiya M.K		
		ICICN16-CS111	A Framework On Penetentiary Conservancy	SyedaAsraSamreen Bhanu K. N.		
		ICICN16-CS106	Implementation of Inpainting Algorithm.	Rohan Ramakrishna Genne Bhanu K. N.		
		ICICN16-CS190	Selective Harmonic Minimization method For Cascaded Multilevel Inverters using ANNs	PrakruthiVasanth Gowtham N		
		ICICN16-CS221	Analyzing The Real Photo Images With The Scanned Photo Images Using Histogram Equalizer	Dr.M.Malathy, Vijayanand, Dr.Arputha Vijaya Selvi		

Programme Schedule | Day – 2 | Theme: NETWORKING & CLOUD COMPUTING **Venue: CSE Seminar Hall**

Date	Sessions/ Timings/	Paper ID	Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 4:- 1.45 pm – 3.15pm	ICICN16-CS187	On Survivability Testing and Computing the Node Connectivity of a Topological Structure of a Interconnection Network Graph	Nethravathi B Kamalesh V N	**Prof. Sreepathi** RRCE, Bengaluru	**Dr. R. Balakrishna** RRCE, Bengaluru
		ICICN16-CS192	Secure System With Higher Reliability And Confidentiality Using Distributed Deduplication Of Multimedia	AishwaryaR,Anusha G Bedekar, Chandana R , Dr.Malathy K		
		ICICN16-CS193	Efficient Schemes for Resource Allocation In the Cloud for Media Streaming Applications	Ms.Janaki ,ShreyaKulkarni PragnaPrakash.S ,Shaila.S		
		ICICN16-CS195	An overview of Swarm Intelligence based Algorithm for Optimization Problem in Wireless Sensor Networks	Nandini G Dr J Anitha		
		ICICN16-CS196	A Scheduled Based MAC Protocols for Wireless Sensor Network:A Survey	Anitha K ,Usha S		
		ICICN16-CS201	NFC-Based Secure Mobile Healthcare System	Madhura P M ,Palash Jain Harini Shankar		
		ICICN16-CS208	A Reward-based MAC layer scheduling For Co-operative Multi-Hop Cognitive Radio Networks	HaseebaYaseen , SnigdhaSen		
		ICICN16-CS209	Li-Fi Based Audio Communication and Device Switching	Madan Kumar K, Karthik M C, Manoj Kumar C, Harshith C, D Pradeepa, K R Pavan Kumar		
		ICICN16-CS218	Performance Analysis Of Various Manet Routing Protocols Using Opnet Simulator	D.Chitra.1 and Dr.M.Chandrasekaran		
		ICICN16-CS220	Detection and Identification of LBBB and RBBB Rhythms in ECG Waves using Gabor Transform Analysis	Geetha A PT. , Asharani M R Gopalakrishnan Nair		
		ICICN16-CS212	Performance Analysis Of Different Filters With Various Noises In Preprocessing Of Images	Janaki K,Dr.M.Madheswaran		

Programme Schedule | Day – 2 | Theme: NETWORKING & CLOUD COMPUTING Venue: CSE Smart Room (No. 522)

Date	Sessions/ Timings/ Paper ID		Paper Title	Author(s)	Session Chair	
13th MAY 2016	Session – 4:- 1.45 pm – 3.15pm	ICICN16-CS182	Be-Safe: Application To Limit Druken Drive Cases	Bharath B K Abdul Rahoof K Ashwin Suresh ,K Rahul	**Dr. Subhashini K** RRCE, Bengaluru	**Dr. Selva Rani** Alliance, University, Bengaluru
		ICICN16-CS215	Wireless robot control with robotic arm using mems and zigbee	Rakesh. H .K		
		ICICN16-CS219	Offline Handwriting Identification Using Adaptive Neural Fuzzy Inference System	Devi T, Swathi Priya N, Prabaharan J		
		ICICN16-CS117	Implementaion Of Secure Data In Cloudcomputing By Double Encryption Method	Govindaswamy H R Bhanu K. N.		
		ICICN16-CS130	Enhancing Computer Inspection Using Document Clustering for Analysis	Mr.Chandrasekhar M S Mr.Subhash B N , Mr.Satish V Mr.Bharath J		
		ICICN16-CS102	Intelligent Software Agent based Image Fusion in Wireless Multimedia Sensor Network	Bhanu K. N Prof.Dr.P.K.Srimani		
		ICICN16-CS204	A Survey on Human Activity Recognition using Android Smartphone.	Usharani J Dr. UshaSakthivel		
		ICICN16-CS217	An Autonomous Robot For Finding Optimal Path During Dynamic Environment Using Grid Based Approach	Jashwanth N B SreeRajendra		
		ICICN16-CS090	Smart Bus Alert System For Easy Navigation Of Blind	Adarsh Holikatti Dr. S. Mohan Kumar		
		ICICN16-CS118	An A3p Approach Towards Image Privacy Policy ecommendation On Content Sharing Sites	MRS. Swapna R DR. Nagesh H R		
		ICICN16-CS120	Data Protection In The Cloud By Using Jar Generation	Hanusha.K ,G.Priyanka		



EDITORS: Dr.T.R.Gopalakrishna Nair, Dr.Usha S, Dr.Malathy, Mr.Rajesh K.S, Dr.Subhashini

PREFACE

ICICN-16 is the 1st International Conference on Innovations in Computing and Networking. The aim of ICICN-16 is to provide a platform for world research leaders and practitioners, to discuss the full spectrum of current theoretical developments, emerging technologies, and innovative applications in Computing, Networking and emerging technologies. Computing and Networking is currently one of the most exciting research areas, and it is continuously demonstrating exceptional strength in solving complex problems. The main driving force of the conference is to further explore the fascinating potential of Computing and Networking.

ICICN-16 includes Keynote addresses, Pre conference Tutorials and Panel Discussion. The time and effort of all the distinguished Keynote speakers, Program committee members and eminent persons from industry handling tutorials are greatly acknowledged.

Keynote speaker in ICICN-16 is Prof. Sisir Roy (Chair Professor, National Institute of Advanced Studies, Bengaluru, Former Professor, Indian Statistical Institute, Kolkata).

Tutorials demonstration in ICICN-16 include, "Software Engineering for service and cloud computing and Enterprise Architecture for Large scale Reuse " by Prof. M. Ramachandran (Leeds Becket University, UK), "Introduction to Topological Quantum Field Theories" by Prof. John Bryden (PM University, Saudi Arabia), "Multi-Media Analytics" by Dr. Om D Deshmukh (Senior Research Scientist, Area Manager, Multi-Media Analytics Xerox Center, India) and "Cloud and its Applications" by Mr. Guru Raghaventhiran R (Practice Manager, KPMG, Bengaluru).

Also ICICN-16 includes Panel Discussion on "IOT- Challenges and Proliferation" by Dr. T.R. Gopalakrishnan Nair(Rector, RRGI, Bengaluru), Prof. John Bryden(PM University, Saudi Arabia), Prof. Ramachandran (Leeds Becket University, UK), Dr. Pethuru Raj (Infrastruture Architect, IBM Global Cloud Center of Excellence, IBM INDIA, Bengaluru) and Dr. Sathyan (Head, Dept of Physics, RRCE, Bengaluru.)

All the accepted papers are reviewed by the panel of reviewers. On behalf of organising committee, I express my sincere gratitude and appreciation to our sponsors. We would like thank members of Advisory committee, Technical committee, Review Committee and Organising Committee for their time and efforts in organizing this conference

On behalf of ICICN -16, a special thanks to our Teaching and Non-teaching faculty, Research Scholars, Students of Department of Computer Science and Engineering and all others for their support throughout the conference. We thank all the authors for submitting their research work to ICICN -16. I would like to thank all the people involved in organizing this conference.

Dr. Usha. S
Organizing Secretary
ICICN - 16

Chairman's Message

"We cannot always build the future for our youth, but we can build our youth for the future."

I am very happy and proud that RajaRajeswari College of Engineering is hosting an International Conference of this stature and at the outset; I congratulate all those who have been a part of this conference. Such conferences give platform for Intellectual exchange wherein experts from the domain come and share their valuable expertise. I have absolutely no doubt that the students, faculties, participants will benefit immensely. Once again my heartiest wishes for all those who have been a part of this conference.

SHRI. A. C. SHANMUGAM

CHAIRMAN

Vice-Chairman's Message

My heartiest congratulations for the team of RRCE for pulling off this conference and hope that the students and the faculties shall be enriched by this experience.

SHRI. A. C. S. ARUN KUMAR

VICE-CHAIRMAN

Executive Director's Message

Greetings!!! Keeping up with today's evolving educational challenges, it requires out of the box approach to technical education as there are lot of technology breakthroughs happening in the educational field today. Professional Institutions need to redesign the training for the students as well as for the faculties. We believe that there has to be a continuous learning curve. Symposiums, workshops, seminars, conferences are the best medium for such scholarly exchanges.

I would like to take this opportunity to express my sincere happiness to all my team who have shown tremendous commitment and enthusiasm in making this happen .

I once again extend a very warm welcome to you all and assure you all a very productive and pleasant time at our institution. Thank you

SHRI. S. VIJAYANAND

Executive Director

Rector's Message

I am immensely happy on the occasion of the commencement of the 1st International Conference on Innovations in Computing and Networking ICICN-16 to be held in RRCE, Bengaluru Soon. I understand that the conference emphasizes on Innovations happened in the recent past in the broad spectrum of computing and networking. As the main objective of ICICN-16 is to provide an elegant platform for researchers to express their novel ideas and practices, this conference is going to be one of the very successful conglomerations of professionals, Encouraging and enriching the computing domain.

Conference received more than 230 papers and it was subjected to sufficient reviews resulting in a selection of 184 papers to be published in the name of the conference. This is a remarkable success. At this juncture, I would like to congratulate all the members of the conference team as well as the philanthropic management who had perceived and realized such an event in Bengaluru.

As the program chair I express my best wishes to all the delegates to a fruitful professional engagement during the period of conference.

Dr. T R Gopala Krishnan Nair

RECTOR

Principal's Message

"Creativity leads to thinking,

Thinking provides knowledge.

Knowledge makes you great"

I, on behalf the organizing committee, very much pleased to invite engineers, technocrats and the teaching faculty to contribute technical papers to International Conference On "1st International Conference on Innovations in Computing and Networking (ICICN-16)", It is designed to attract as many experts from the Computer Science and Engineering from all over the nation to showcase their research and interact for the benefit of all academicians and the industry. I hope that each participant gets individually enriched in turn enriching every other participant. I am sure that this conference accords the right forum for interaction and benefits everyone concerned.

Join us and enjoy yourself at this great event and partake of our hospitality.

As General Chairperson of the ICICN-16, I believe that this conference will help us in many directions to develop our country research potential.

Dr.R.BALAKRISHNA

Principal

Vice Principal's Message

Proud to know that the department of Computer Science & Engineering of our Institution is conducting 2 days International Conference on 12,13 May 2016 and pre conference tutorial on 11 May 2016. The conference will probably chase for new innovations.

I am optimistic that, at the end of the day, many more innovative ideas on computing and networking could be place.

My congratulations to organizers and best wishes for the event.

Prof. PRABHAKAR M

Vice Principal

Organizing Secretary Message

The department of Computer Science & Engineering, RajaRajeswari College Of Engineering, Bengaluru proudly organizing our 1st International Conference "Innovations In Computing and Networking(ICICN16)" on 11,12,13 May 2016.

On behalf of the organizing committee ICICN16, I proudly welcome all the speakers and delegates of different parts of the country and the world to present their innovative ideas on current trends in technology.

I would like to thank the keynote speaker and all the tutorial speakers. I will also place my gratitude to the management of RajaRajeswari group of Institutions. I would also like to thank all the members of different committees, who worked successively to make this event in much bigger way.

I wish a good luck for all the participants.

Dr. USHA S

Organizing Secretary

Convener Message

The conference is a meeting and information exchange between the end user, the event has attracted academicians, research scholars, students and industry experts. This conference will be exceptional in many ways, a strong accentuate is being prepared for a well-balanced and effective academic - industrial participation with High quality papers and presentations ; a momentous number of Keynotes, Invited talks and presentation sessions delivered by delegates from academic and industry communities.

I hope that this conference would surely induce modern ideas among the participants paving way for new inventions in the field of computer science

We are also very much thankful to the Management, Rector and Principal

I would like to pay thanks to all Authors and participant for their association.

Prof. RAJESH.K.S

Convener

Convener Message

It is a personal honor and pleasure for me to be the Convener of ICICN16. The technology today is developing at a rapid pace. In this era of globalization, the exchange of knowledge and skills has given a further fillip to the exponential growth in the field of technology. Such phenomenal advancements have revolutionized almost every sphere of human life today.

The objective of this conference is to provide a concrete platform which will encourage & support scholars, researchers & industry professionals to carry & accomplish their research targets.

Expecting a sound response from you all.

Prof. SRINIVAS.R

Convener

CHIEF PATRON

Sri.A.C. Shanmugam
Chairman,
Moogambigai Charitable Educational Trust,
RajaRajeswari Group of Institutions, Bengaluru

PATRON

Sri. A. C. S. Arunkumar
Vice Chairman,
RajaRajeswari Group of Institutions, Bengaluru

GENERAL CHAIRS

Sri. S Vijayanand , Executive Director, RRGI
Sri. Jeyabalan , Special Officer, RRGI
Dr. R Balakrishna , Principal, RRCE

PROGRAMME CHAIR

Dr. T R Gopalakrishna Nair, Rector, RRGI

ORGANIZING SECRETARY

Dr. Usha S, Prof. & Head, CSE, RRCE

CONVENORS

Prof. Rajesh K S
Prof. Srinivasa R

Co-CONVENORS

Prof. Sreenivasa B R
Prof. Aruna kumara B
Prof. Swetha P

ADVISORY COMMITTEE

Dr. S S Iyengar, Prof., Louisiana State University, Louisiana.
Mr. Arvind S ,Senior Engineer, Intel Corporation, USA.
Dr. M Ramachandran, Leeds Beckett University, UK.
Dr. Hong Jer long, Prof., Taylor's University, Malaysia.
Dr. John Bridan, PM University, Saudi Arabia.
Dr. S R S Iyengar, Prof., IIT Ropar, Chandigarh.
Mr. Sukumar, Tech Specialist, RBS Development Center, Chennai.
Dr. Daniel Chandran, Dy. Head, University of Techno, Sydney, Australia.
Dr. Pradeep B S, Director, R&D Dept., LinyiTop N/W, Shandong, China.
Dr. Lim Tien Sze, Prof., Multimedia University, Malaysia.
Dr. Maulik Ujjwal, Jadavpur University, Kolkatta.
Dr. Cyril Raj, Dean S&T, Dr MGR University, Chennai.
Mr. E Y A Charles, Prof., University of Jaffna, Srilanka.
Dr. Niranjan N C, Principal, NMAMIT, Udupi.
Dr. Umakanth P Kulkarni, Prof., SDMCE, Dharwad.
Mr. Chandrappa Upparahalli, General Manager, RMHS Dept., JSW, Bellary.
Mr. Rajasekar, Senior Manager (Systems) , Afcons, Kuwait.
Dr. Arokiaswami Alphones, Assoc. Prof., NTU, Singapore.
Dr. S Radha, Prof.& Head, SSNCE, Chennai.
Mr. A. Ramesh, Senior system Analyst, Cadence Systems, Bengaluru.
Mr. Paul Raj, GE, Bengaluru.
Dr. Shadaksharappa Bichagal, Prof., & Head, CS&E, SSEC, Bengaluru.
Dr. S Raghavan, Prof., NIT, Trichy.
Dr. Nirmala C R, Prof .& Head, BIET, Davangere.
Mr Ram GurusamyRaja, Application Development Analyst,Reciter Affiliated Companies,Oxnard | CA,USA
Dr. Om.D Deshmukh, Senior Research Scientist, Multimedia analytics, Xerox Research Centre, Bengaluru.

PROGRAM COMMITTEE

Dr. L M Patnaik, Prof., INSA, FNAE.
Dr. M. Arunachalam, Prof.& Head, EEE, FNAE, RRCE, Bengaluru.
Dr. Guru Prasad, Prof.& Head, CSE, BMSCE, Bengaluru.
Dr. Nithya G Nair, University of Manchester, UK.
Dr. Usha, Prof., RVCE, Bengaluru.
Dr. Santhosh Nair, Co-founder, AngelEX, Bengaluru.
Dr. Usha S, Prof. & Head, CSE, RRCE, Bengaluru.
Dr. Arun Biradar, Prof.&Head, CSE, EWIT, Bengaluru.
Dr. Meenakshi Sundaram, Prof.&Head, CSE, DBIT, Bengaluru.
Prof. Prabhakara.M ,Vice Principal ,RRCE, Bengaluru.
Dr. Asha, Prof., BIT, Bengaluru.
Dr. Guruprasad N, Prof., New Horizon, Bengaluru.
Dr. Latha C A, Prof., CSE, GAT, Bengaluru.
Prof. Auntin Jose, Assoc. Prof., CSE, RRCE, Bengaluru.
Dr. Sheela T, Prof.,&Head, IT, SSEC, Chennai.
Dr. Pillai C S, Prof.&Head, CSE, ACSCE, Bengaluru.
Dr. P V Rao, Prof., ECE, RRCE, Bengaluru.
Dr. T. Senthil Kumaran, Prof., ACSCE, Bengaluru.
Mr. Samuel Raja Singh, Senior Manager, Flextronics India PVT. Ltd., Bengaluru.
Mr. M Guru Ragaventhiran R, Practice Manager, Management consulting, KPMG, Bengaluru.
Mr. Shreyas Shivaprakash, Web Operations Analyst, Best Buy, USA.
Dr Subhashini K, Prof., CSE, RRCE, Bengaluru.
Mr. Girish Manchaiah, E-Commerce Architect, Oracle Corporation, USA.
Mr. A.V. JoshiKumar, Senior Engineer, Accenture, Singapore.
Dr. K Maran, Director, SIMS, Chennai.
Dr. Subburam, Prof., PSVPEC, Chennai.

REVIEW COMMITTEE

Dr. M Ramachandran, Leeds Beckett University, UK
Dr. Nithya G Nair, University of Manchester, UK.

Dr. R .Selva Rani, Chair Person, CSE, Alliance University, Bengaluru.
Dr. G. T. Raju, Prof.& Head, CSE, RNSIT, Bengaluru.
Dr. R Balakrishna , Principal, RRCE.
Dr. SwarnaJyothi, Prof.&Head, ECE, RRCE, Bengaluru.
Dr. JayaPrakash, Prof.&Head, ISE, BIT, Bengaluru.
Dr. Santhosh Nair, Co-founder, AngelEX, Bengaluru.
Dr. Kavitha Soode, Prof., NMIT, Bengaluru.
Dr. Marry Cherian,Prof., AIT, Bengaluru.
Dr. Sheela T, Prof., & Head, IT, SSEC, Chennai.

INAUGURATION AND VALEDICTORY COMMITTEE

Mr. Rajesh K S
Ms.Poonam Kumari
Dr.Malathy M
Mr.Shashidhar V
Ms.Swetha P
Ms.Anitha K

SESSION COORDINATORS COMMITTEE

Dr.Malathy M
Mr.Neelu L
Ms.Janaki K
Ms.Vandana B
Mrs.Nandini G

TUTORIALS COMMITTEE

Mr .Nandini.G
Mr. Bharth.J
Mrs.Anitha K
Mr. Aravind

TRANSPORTATION COMMITTEE

Mr.Aruna kumara B
Mr.Kamal Raj T
Mr.Bharath J
Mr.Manigandan J

Mr.Srinivas.R

CERTIFICATE COMMITTEE

Ms.Poonam Kumari
Dr.Malathy
Ms.Neelu L
Ms.Vandana B

HOSPITALITY COMMITTEE

Ms.Poonam Kumari
Mr.SarvanaPerumal.
Mr.Kamal Raj T
Mrs.Nandini G
Ms.Neelu L

RECEPTION COMMITTEE

Ms.Poonam Kumari
Mrs.SwathiPriya. N
Ms.Swetha P

REGISTRATION COMMITTEE

Mrs.Anitha.K
Mr.Sreenivasa. B.R
Mr.Shashidhar V

STAGE COMMITTEE

Mr. Srinivasa R
Mr.Auntin Jose
Mrs.Nanadini G
Mr.Prabhaharan
Mr.Saravan Perumal V M
Mr.Kamal Raj T

REFRESHMENT COMMITTEE

Mr. M H Rehaman
Ms.Anitha K
Mr.Manigandan J
Mr.Prabaharan J

JOURNALS AND PROCEEDINGS COMMITTEE

Dr.Usha S
Mr.Rajesh K S
Mr.Shashidhar V
Dr.Malathy M
Ms.Janaki M
Ms. Neelu L
Mr.Sreenivasa B R

EDITORIAL COMMITTEE

Dr.T.R.Gopalakrishnan Nair
Dr.Usha S
Dr.Malathy
Mr.Rajesh K.S
Dr.Subhashini

COMMUNICATION AND ADVERTISEMENT COMMITTEE

Ms.Swetha P
Ms.Sreenivas.B.R
Mr.Manigandan J
Mr.Aruna Kumara B

SPONSORSHIP COMMITEE

Ms.Neelu L
Mr.Bharath.J
Mr.Saravanaperumal
Mr.Kamal Raj T

WEBSITE COMMITTEE

Dr.Usha S
Mrs. Swetha P

FINANCE COMMITTEE

Mr. T. AuntinJose
Ms. Vandana
Mr. Md.Rehman

PURCHASE COMMITTEE

Mr. Sreenivasa.B.R
Ms. M H Rehaman
Mr. Saravana Perumal
Mr. Aruna kumara B

BOARDING AND LODGING COMMITTEE

Mr. Bharath J
Mr. SarvanaPerumal. VM
Mr. Srinivas.R
Mr. Aruna kumara B

INDEX

Improving the Wireless Sensor Network Lifetime by Diminishing the Energy Consumption via Nature Inspired Technique

Vinay R N, Parvathi C
Department of Computer Science, Visvesvaraya Technological University, Karnataka
rnvinay.1992@hotmail.com, pvc3119@gmail.com

ABSTRACT: Wireless detector Networks (WSN's) is one among the foremost leading technologies round the globe that build it attainable the embedded systems of case size with the similar skill of a private laptop. It consists of autonomous battery battery-powered devices referred to as nodes that act as an electrical device. The nodes area unit intelligent enough to sense, calculate and send the info. Because the nodes area unit battery battery-powered instrumentality, therefore one among the crucial task of the detector network is to manage the use of energy. The area unit various aspects that cause associate energy deficiency within the network. One among them is energy exhaustion whereas transmittal information as a result of the energy absorbed in transmittal the info is double the energy used in reworking the info. Hence, the transmission route ought to be such that drains lowest energy together with productive dispatching of knowledge. The paper focuses on this angle and proposes a nature galvanized approach to accumulate a route of high energy for addressing information to the destination to diminish the energy consumption and rise the network period.

Keywords - WSN, ACO, MPRSO, Routing

I. INTRODUCTION

The Sensing submissions has created the time period of users so swish and accelerated. Nowadays, might access the big machinery whereas not pushing aside our hands from the pocket similar to the escalators begin moving, the time amendment front of it. All applications like this area unit potential with the evolution of wireless device networks. A Wireless device Network is associate integration of the nodes that execute their task during a very collective manner. These wireless nodes can accomplish the operate of sensing, aggregating and method the data whereas maintaining relevance completely different nodes of the network via oftenest channels. WSN's area unit loosely utilized in an exceedingly vary of applications like military, medicals, defense, environmental, etc. Basically, device nodes contain units and these area unit of four types: sensing unit, method unit, transceiver unit and unit of measurement, which is shown in fig 1

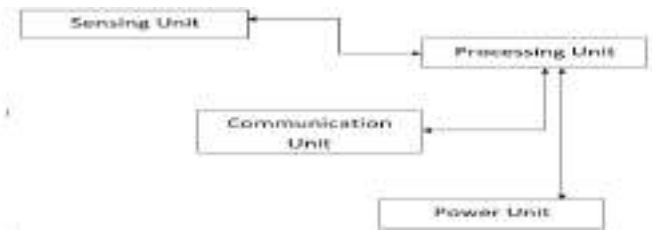


Fig 1: Wireless Sensor Node Architecture

Sensing unit is any of classified in to a pair of subunits i.e. sensors and AD converters, physical phenomena (like temperature, sound, light, image) into electrical signals. The analog signals that square measure created by sensors square measure reworked into digital signals by practice analog to digital convertor before causing to method unit. Processing unit contains chip that controls the devices and performs execution of communication protocols in conjunction with the signal method algorithms on the collected device information. Transceiver unit is in charge of creating a communication wirelessly with neighboring nodes and conjointly the skin world and consists of electrical battery that offers power to the complete device network.

Energy consumption is one altogether the foremost causes, that limits the performance of the network. It's getting to get minimized among the action of transmission, reception or idle listening or overhearing. The route formation for transmission information packets is one altogether the solution to beat this downside as short routes that consists of nodes with exhausted batteries would possibly ends up in poor network amount whereas on the other end long routes would possibly contains huge vary of nodes which can raise the network delay.

Choosing a quick route would possibly ends up in depletion of intermediate nodes that lands up into decrease in amount of the network, but at an identical instant shortest route may contributes good lands up in terms of low energy consumption and network amount. thus numerous routing algorithms square measure there among the literature, variety of them square measure nature affected means the Bio-inspired optimization algorithms notably insect Colony optimization (ACO), Genetic Algorithms (GA) and Particle Swarm optimization (PSO) etc. The ACO formula depends on the behavior of ants searching for food in an exceedingly neighborhood.

The GA is associate optimization methodology that depends on the Darwin's principle of biological enlargement, duplicate and conjointly the survival of the fittest. The PSO formula depends on the collective behavior of birds or swarms. This may be a population primarily based search methodology throughout that particle changes its position with time. Throughout this procedure, each particle uses its own experience and conjointly the experience of its neighbor to manage its best position. But still there is analysis on minimizing the energy consumption and decreasing the node ruptures. Being affected by these algorithms, planned associate formula which could modify the current Particle Swarm optimization (PSO) formula

throughout that alone the position update is utilized. The simulation area unit getting to be applied in NS2 and conjointly the results of the planned technique square measure compared with the current routing algorithms.

II. RELATED WORK

Research Methodology

A clump based mostly routing algorithmic rule named as Low Energy accommodative clump Hierarchy (LEACH) has been developed. LEACH forms the clusters of the nodes during a distributed manner and chooses a 1 node as a cluster head on the idea of sure likelihood. The method is separated into rounds. Every section begins with accelerate section, that collects and aggregates the information and followed by a gradual state section that holds the responsibility to transmit the information.

A pismire Colony improvement (ACO) algorithmic rule is employed that forms clusters by selecting the node containing most residual energy than average energy price as a cluster head in order that it will overcome the first death of the node. In every cluster to take care of balance among the nodes repetitive division methodology is employed.

The Particle Swarm improvement (PSO) algorithmic rule is employed to optimize a path in Wireless detector Networks. This algorithmic rule finds out that improvement of the routing in WSN is superior in terms of quality and high success rate is achieved as compared to the Genetic algorithmic rule.

Particle swarm improvement with random search improvement techniques referred to as MRPSO for the answer. This MRPSO uses solely position update, whereas rate update is avoided.

III. PROBLEM FORMULATION

Although the wireless device networks are operated during a vast range of applications, however they conjointly face some constraints too, for instance confined transmission vary of device node, restricted process and storage potentials similarly as their energy resources. Because of finite energy resource the most task of the wireless system is to execute transmission fruitfully by maintaining the energy. For this a route formation is important by choosing the energy economical nodes. The nodes of the network keep some coordination to reinforce the performance of the network.

Some things throughout packet transmission, the nodes could get dead as a result of depleted energy resources, or any quite physical harm, failure within the software system or could get destroyed because of environmental modification which ends in touching the network period of time. to beat from this downside developed a way to seek out an appropriate route containing higher energy resource supported nature primarily based algorithms to transfer a packet, leading to minimum energy consumption and enlargement of the network period of time.

IV. PROPOSED TECHNIQUE

Proposed technique the sensing area is divided into cells and in each cell the equal number of nodes are deployed and initial energy values are assigned to each node. Communication would occur from source to destination and it is mandatory that the route gets into each cell of the network and selects one node from each cell. First of all the initial population is generated which randomly gives the set of routes between source to destination.

The total energy of each route is calculated and the maximum energy value obtained by any route is considered as a best fitness value. This is the initial solution. To obtain a route the iterations are carried out using MRPSO which helps to update the initial population. The total energy value of each updated route is calculated, if the energy of the updated route is higher than the best energy value obtained from the initial population, i.e. initial solution then accept it else it moves to the next iteration. Updating the initial population would be obtaining the Pbest (Position Best) solutions and all the Pbest solutions are further considered to get the gbest which is the route containing the maximum energy value among all the Pbest solution. Finally, as per the gbest (Global Best) the energy efficient route is achieved.

The methods used for planning the algorithm are defined as follows along with the flow chart of planned algorithm.

A. Algorithms

Ant colony optimization (ACO method)

Ant colony algorithms square measure supported the behavior of ants finding food in a very search space. This rule is employed for determinant the best methods from supply to food. For the primary time hymenopterous insect roams willy-nilly, once they found food ants come back to their colony and marked their followed path by pheromones that shows that the trail has food. Once alternative ants see these markers of secretion they have a tendency to travel when constant path with some bound likelihood so as to bring food. The trail gets inhabited with their own secretion and also the path would get stronger as several as ants follow constant path. Shortest methods square measure stronger than longest ones as a result of in shortest path the massive quantity secretion is gift, whereas in longest one it should get decayed.

Particle swarm optimization (PSO Method)

PSO is Associate in Nursing improvement formula supported bird flocking looking for the food in Associate in Nursing exceedingly} terribly search space. Every particle noted as swarm gathers the data from every array build up by their various positions. Modification of particle's positions is completed by pattern the speed of the particle. Particle's own expertise so the expertise of its neighbors is used to update the position and rate of a particle. Thus on modification the worldwide search ability of the PSO use a MRPSO formula that uses solely position vector and no rate vector is used

Updated position = bestfitnessvalue + αβ (mbestcurrent position)

Where, mbest = pbest / population size, α is 0.37 and β= (rand1-rand2)/rand3.

The rand1, rand2 and rand3 contain value within 0 to 1.

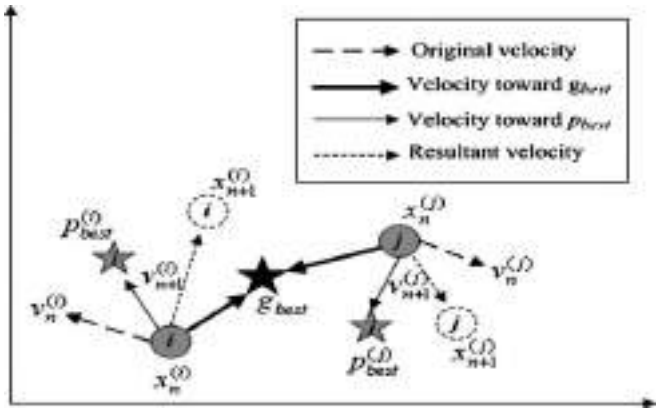


Fig 2: PSO working

B. Flowchart

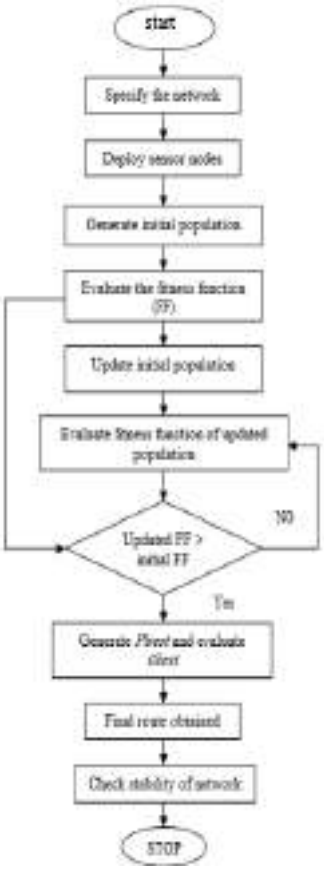


Fig. 3.Diagram for the Proposed Algorithm

V. PERFORMANCE EVALUATION

This experiment thought-about a sensing space that area unit divided into nine blocks such the sensing region have three rows and three columns of equal size. Then equal range of nodes area unit deployed in every block. Block one is appointed to supply and block nine is appointed to the destination. Further, the node S2 is assumed as a supply node and therefore the node D9 is assumed as a destination node. Size of initial population is ready to ten. Energy consumption for transmitter and receiver area unit set to fifty nJ/bit. The energy consumption issue for the free house and multipath is ready to ten pJ/bit/m2 and zero.0013 pJ/bit/m4 severally.

The energy wont to combination the info is ready to 5nJ/bit/signal. Information packet size is ready to 600bytes. {The range the amount the quantity} of alive nodes is calculated with relevance number of rounds. Once the LEACH operates the primary node dies at the spherical sixty and within the case of ACR the primary node dies at the spherical 157 whose performance is healthier than the LEACH.

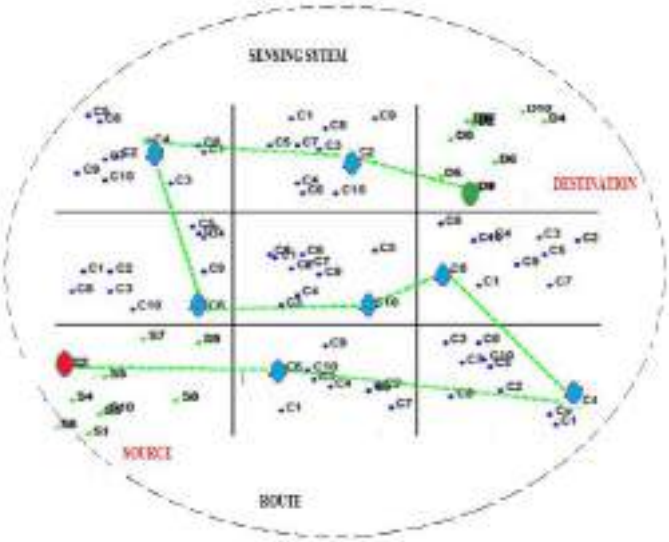


Fig.4. Final optimized Route.

VI. EXPERIMENTAL RESULTS

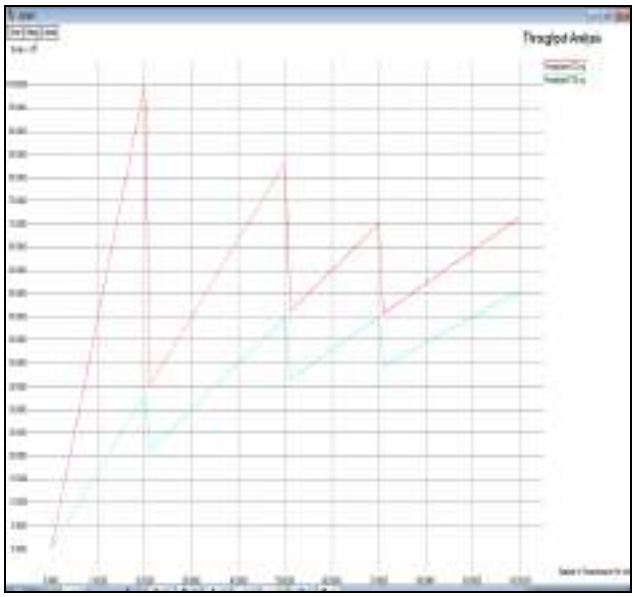


Fig.5. Throughput Analysis of ACO &PSO

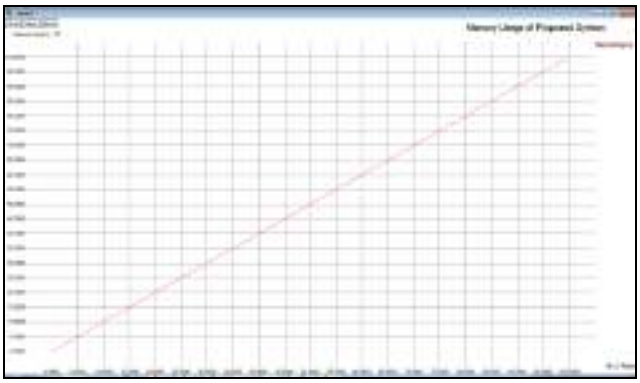


Fig.6. Packet Memory Usage of Proposed System

VII. CONCLUSION

An energy economical routing could be an important issue within the wireless detector network. During this work a nature impressed energy economical routing algorithmic program is meant and simulated. It determines the route for knowledge transmission from supply to destination by victimization the improvement algorithmic program. Projected technique outperforms over Low Energy adaptive clump Hierarchy (LEACH) and hymenopter on Colony Routing (ACR) because the 1st node of those algorithmic program dies early than technique. The first node dies get into projected technique is at 173th spherical, whereas within the case of LEACH the primary node dies out at sixtieth spherical and in ACR the primary node dies out at 157th spherical, which suggests the present technique consumes additional energy as nodes dies before than the projected technique.

VIII. FUTURE ENHANCEMENTS

In future we have a tendency to concentrate a lot of on ACO formula to enhance its potency by adding bound attributes over it like

- Node Strength enhancements,
- Positions Concern of every node and
- Enhancing the Attack Detection Procedures.

So that the long run formula is finer compare to the projected methodology

REFERENCES

Journal Papers:

[1] J. Yick; B. Mukherjee; D. Ghosal, "Wireless sensor network survey", Comput. Networks, Elsevier, *vol. 52, no. 12, pp.2292 -2330, 2008.*

[2] Akyildiz, I.F.; Weilian Su; Sankarasubramaniam, Y.; Cayirci, E., "A survey on sensor networks," Communications Magazine, *IEEE, vol.40, no.8, pp.102-114, Aug 2002.*

[3] Lindsey, S.; Raghavendra, C.; Sivalingam, K.M., "Data gathering algorithms in sensor networks using energy metrics," Parallel and Distributed Systems, *IEEE Transactions on, vol.13, no.9, pp.924-935, Sep. 2002.*

[4] Di Caro, Gianni; Ducatelle, F.; Gambardella, L.M., "Swarm intelligence for routing in mobile ad hoc networks," Swarm Intelligence Symposium, *2005. SIS 2005. Proceedings 2005 IEEE, pp.76-83, 8-10 June 2005.*

Theses:

[5] Dorigo, M.; Birattari, M.; Stutzle, T., "*Ant colony optimization*," Computational Intelligence Magazine, IEEE, vol. 1, no.4, pp.28-39, Nov. 2006.

[6] Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H., "*Energy-efficient communication protocol for wireless microsensor networks,*" System Sciences, Proceedings of the 33rd Annual Hawaii International Conference on, pp.10-14, 4-7 Jan. 2000.

[7] Yulong Shen; Qingqi Pei; Qijian Xu; Hailin Feng; Jianfeng Ma, "*A Routing Algorithm Based on Ant-Colony in Wireless Sensor Networks*," Computational Intelligence and Security, CIS '09, International Conference on, pp. 441-445, 11-14 Dec. 2009

[8] Zhu Xia; Zhang Yulin, "*Wireless sensor network path optimization based on particle swarm algorithm*," Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on, pp.534537, 10-12 June 2011.

[9] Nagendra Singh, and Yogendra Kumar,"*Economic load dispatch with valve point loading effect and generator ramp rate limits constraint using MRPSO,*"International Journal of Advanced Research in Computer Engineering & Technology(IJARCET), Vol. 2, no. 4, pp. 1472-1477, April. 2013.

[10] Lee, J.-W.; Byoung-Suk Choi; Ju-Jang Lee, *"Energy-Efficient Coverage of Wireless Sensor Networks Using Ant Colony Optimization With Three Types of Pheromones*," Industrial Informatics, IEEE Transactions on , vol.7, no.3, pp.419-427, Aug. 2011.

An Analytical Framework Intended For Congestion in Wireless Networks

Daina K K, Dhanalakshmi H.B , Pallavi Gowdoor , Sindhura D
Department of Information Science , SKIT, Bengaluru-90.
dainakariappa@gmail.com, dhanu.hb7@gmail.com, pallavigowdoor12@gmail.com, sdsindusarala@gmail.com

ABSTRACT: Application flows that contains congestion or video traffic is taken into account. Reducing the level of video distortion is essential per user perspective. Common link-quality-based routing metrics (such as ETX) don't account for congestion across the links of a path, as a result, they'll cause video flows to converge onto a couple of methods and, thus, cause high video distortion. We have a tendency to construct associate degree analytical framework to know and assess the impact of the network on video distortion. The framework permits us to formulate a routing policy for minimizing distortion, supported that we have a tendency to design a protocol for routing video traffic. We find based on simulations and testbed experiments that our protocol is economical in reducing video distortion and minimizing the user expertise degradation.

KEYWORDS—Protocol design, routing, video communications, video distortion minimization, wireless networks.

1. INTRODUCTION

With the arrival of smart phones, video traffic has become an extremely popular in wireless networks. per user perspective, maintaining a good quality of the transferred video is crucial. The video quality is affected by: 1) the distortion because of impression at the source, and 2) the distortion because of each wireless channel induced errors and interference. Video cryptography standards, like MPEG-4 [1] or H.264/AVC [2], define groups of I-, P-, and B-type frames that give totally different levels of encoding and, thus, protection against transmission losses. The various levels of cryptography refer to: 1) either data encoded independently, in the case of I-frames, or 2) cryptography relative to the information encoded inside other frames, as is the case for P- and B-frames combining together they form a group of pictures (GOP). Critical functionalities that's usually neglected, however affects the end-to-end quality of a video flow, is routing ancient routing protocols, designed for wireless multihop settings, are application-agnostic. User-perceived video quality will be significantly improved by accounting for application needs, and specifically the video distortion experienced by a flow, end-to-end. The schemes used to encipher video clip can accommodate a particular number of packet loss.

If the amount of lost packets in a frame exceeds a certain threshold, the frame cannot be decoded correctly. A frame loss can end in some quantity of distortion.

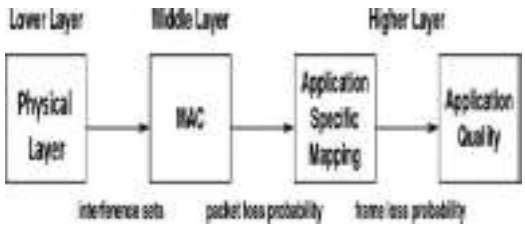


Fig.1. Multilayer approach

As one of our main contributions, we construct an analytical model to account for evolution of frame losses within the GOP as video is delivered on an end-to-end path. In our model, we capture however the selection of path for associate degree end-to-end flow affects the performance of a flow in terms of video distortion.

Our model is constructed based on a multilayer approach as shown in Fig. 1. The packet-loss probability is mapped to the frame loss probability within the GOP. The frame loss probability is then directly related to the video distortion metric. By using the on top of mapping we discover the path from source to destination that minimizes end to end distortion.

Our contributions during this paper are as follows:

• Developing an analytical framework to find the effect of routing on video distortion:

The framework facilitates the computation of routes that are optimum in terms of achieving the minimum distortion.

• Design of a sensible routing protocol for distortion-resilient video delivery:

The practical protocol permits a source to collect distortion data on the links within the network and distribute traffic across the various paths in accordance to: 1) the distortion, and 2) the position of a frame in the GOP.

• Evaluations via extensive experiments:

We show via extensive simulations and real testbed experiments that our protocol is very effective in reducing the end-to-end video distortion and keeping the user expertise degradation to a minimum.

2. RELATED WORK

Standards just like the MPEG-4 [1] and also the H.264/AVC [2] give tips on how a video clip should be encoded for a transmission over a communication system supported layered cryptography. H.264/AVC is newest video cryptography commonplace of the ITU-T Video cryptography experts group and the ISO/IEC picture consultants group. The main goals of the H.264/AVC standardization effort have been increased compression performance and provision of a "network-friendly" video

representation addressing "conversational" (video telephony) and "non-conversational" (storage, broadcast, or streaming) applications. H.264/AVC has achieved a significant improvement in rate distortion efficiency relative to existing standards. This article provides a summary of the technical features of H.264/AVC, describes profiles and applications for the quality, and descriptions the history of the standardization method.

In [3] it presents the expected transmission count metric (ETX), that finds high-throughput paths on multi-hop wireless networks. The ETX metric incorporates the consequences of link loss ratios, asymmetry within the loss ratios between the 2 directions of every link, and interference among the successive links of a path. In distinction, the minimum hop-count metric chooses at random among the different paths of a similar minimum length, regardless of the usually massive variations in throughput among those paths, and ignoring the possibility that an extended path might offer higher throughput.

In [4], an analytical framework is developed to model the effects of wireless channel fading on video distortion. We propose an accurate and fully analytical model for the distortion due to lost packet in wireless communication system. The model is, however, only valid for single-hop communication.

In [5] and [6], MDC is considered for video multicast in wireless ad hoc networks. Although these 2 papers think about the distortion since they're using MDC technique. Our approach differs not only on the way we model video distortion, however additionally on the actual fact that we tend to target LC, which is a lot of popular in applications nowadays. The Multiple Description cryptography (MDC) technique fragments the initial video clip into a number of sub-streams known as descriptions. The descriptions are transmitted on the network over disjoint paths. Even one description received at the receiver entire video is decoded however only if all the n received quality is nice. Layered coding (LC) produces a base layer and multiple enhancement layers. The improvement layers serve only to refine the base-layer quality and aren't useful on their own. Layered coding is employed because of its popularity in applications and adoption in standards. The work in [7] proposes a theme for energy-efficient video communications with minimum QoS degradation for LC. The routing scheme is predicated on a hierarchical model. To support such a hierarchy, the nodes need to be grouped in clusters, and a method of electing a cluster head needs to be executed periodically, increasing the process and data communication load of the network. In our projected theme we have a tendency to assume a flat model where all nodes within the network are equivalent and perform a similar set of tasks.

3. IMPLEMENTATION

3.1. Model Formulation

Our analytical model couples the functionality of the physical and mac layers of network with the application layer for a video that is sent from a source to a destination node. The model for the lower layers evaluates the packet loss probability through a collection of equations that characterize multiuser interference, physical path conditions, and traffic rates between source– destination pairs within the network. This packet-loss probability is then input to a model 2 to compute the frame-loss probability and from that the related distortion. The value of the distortion at a hop on the path from the source to the destination node depends on the position of the first unrecoverable frame in the GOP.

3.2. Video Distortion Model

Our analysis is predicated on the model for video transmission distortion. The distortion is divided into source distortion and wireless transmission distortion over one hop. Instead of focusing on one hop, we tend to significantly extend the analysis for multihop by developing a model that captures the evolution of the transmission distortion on the links of a route from the source node to the destination node. Assuming that the packet losses in different frames within the GOP are independent events (likely if the fading patterns modification in between), the transition probabilities for the method, can be computed.

3.3.Video Distortion Dynamics

The value or number of the distortion at hop on the path from the source to the destination node depends on the position of the primary unrecoverable frame in the GOP. The value zero indicates that the first (I-frame) is lost, and so the full GOP is unrecoverable. a value larger than zero denotes that the corresponding P -frame is that the initial frame in the GOP that can't be decoded correctly and the value indicates that no frame has been lost therefore , yielding a distortion. The dynamics of the method and therefore of the video distortion depend upon the process.

3.4. Optimum Routing Policy

In this module, our objective is to search out the path that yields the less video transmission distortion between any source and destination. By using the analysis given, we tend to pose the problem as a random optimum management problem where the control is that the choice of future node to be visited at every intermediate node from the source to the destination. We tend to decide that this optimization problem is called as Minimum Distortion Routing (MDR) problem.

MDR problem has the following characteristics:
L 1: MDR satisfies the overlapping property, i.e.,

The problem is divided into smaller problems that retain a similar structure.

L 2: MDR satisfies the optimum substructure property, i.e., the sub-path of an optimum path is optimal for the corresponding sub-problem.

Theorem 1.: The MDR problem is resolvable by dynamic programming.

Proof: An optimization problem is resolved by dynamic programming if the problem satisfies both the overlapping and the optimum substructure properties. The proof is immediate from L's one and 2. In essence, the MDR routing policy distributes the video frames across multiple paths and particularly minimizes the interference experienced by the frames that are at the initial of a GOP). The I-frames are

longer than other frames. Their loss impacts a lot of distortion and therefore these are transmitted on comparatively interference-free paths. The upper protection rendered to I-frames is that the key contributive factor in decreasing the distortion with MDR.

3.5. Protocol Design

To calculate the answer to the MDR problem, data of the entire network is necessary. The answer to the MDR problem will be computed by the source node supported partial information relating to the worldwide state that it gathers. The source node should sample the network during a path discovery method so as to gather information relating to the state of the network. The sampling method includes the estimation of the ETX metric for every wireless link within the network. These estimates give a live of the quality of the links. The estimation method is implemented by tracking the successful broadcasting of problem messages in periodic time intervals.

The protocol uses 2 algorithms.

The ETX estimation evaluated locally within the neighborhood of a node is appended within the Route Request messages at the time of Route Discovery phase. Upon reception of this message at the destination, a Route Reply message is sent back to the source. The source node then will solve the improvement problem by using the data gathered via the sampling method described on top of

Specifically, upon receiving the Route Reply messages, the source node executes steps conferred in formula 1. Next, by invoking algorithm two produces future node within the path. This can be done to consider the mobility of the nodes a difficulty of wireless network.

The flowchart that represent the operation of the source node is shown in Fig. 2(a), while the flowcharts for Associate in Nursing intermediate node and therefore the destination node are delineated in Fig. 2(b).

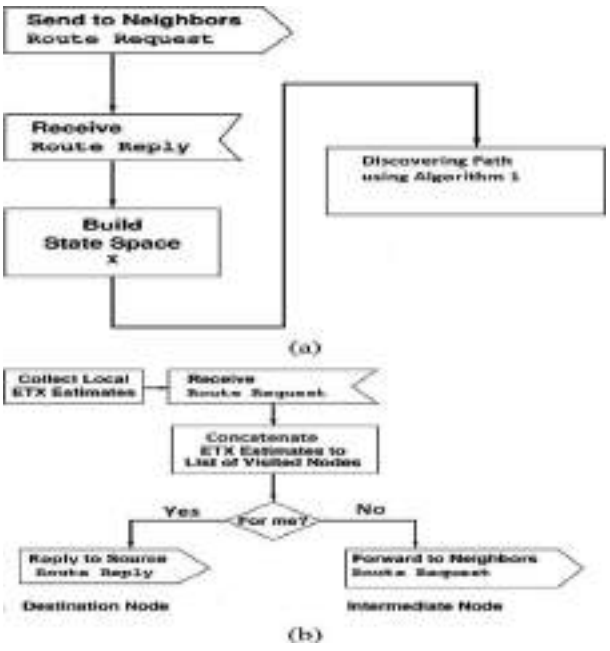


Fig. 2. Flowchart for application -aware routing. (a) Source node. (b) Intermediate and destination node.

4. RESULTS

We represent the performance gains of the proposed routing theme via extensive simulations and testbed experiments. We tend to use the network simulator ns-2. That provides a full protocol stack for a wireless multi-hop network. We tend to extend the functionality of ns-2 by implementing our projected routing theme on top of the present protocol stack. For the testbed experiments, we tend to implement our scheme using the click modular router. We implement 2 totally different strategies and experiment with each, one after another. The first technique estimates the ETX value for every link between a node and its neighbours for all the nodes within the network. The mechanism broadcasts periodically (every one s) little probe messages of size 32bytes and checks for acknowledgments from the neighbours of the node. The routing policy computes the minimum ETX path from the source to a destination and uses that path to transfer the video packets. The second technique implements the protocol defined in implementation so as to compute the routes on the wireless network that achieve minimum video distortion. We tend to use EvalVid which consists of a collection of tools for the analysis of the quality of video that's transmitted over a true or simulated network. The toolset supports totally different performance metrics like the PSNR and also the MOS.

We need to capture a log from an attempted transmission over a true network. This log indicates that frame and at what time instance was transmitted over the network. The log is fed as an input to the ns-2 simulation that plays back the video transmission, producing at the end 2 sets of statistics relating to the transmission, one for the sender and one for the receiver. By applying the EvalVid toolset on this sequence of files, we can reconstruct the video file because it is received by the destination and compare it to the initial video file. The comparison provides a measure of the video quality degradation as a result of the transmissions over the network.

4. Simulation Results

To evaluate the performance of the MDR protocol, we tend to compare it against the minimum ETX routing theme. The pair of nodes that represent the source and destination in every case are selected at random. If they happen to be neighbors, we discard that pair and repeat the method till we tend to select a source and destination that are over one hop apart. Every set of experiments is repeated 10 times, and therefore the average value is reported in every case. Our simulation experiments specialize in 3 metrics: 1) the PSNR, that is an. objective quality measure; 2) the MOS, that is a subjective quality metric; and 3) the Delay experienced by every video connection. The PSNR (Peak signal/noise Ratio) is that the most widespread objective metric to measure the digital video quality, it doesn't always capture user experience.

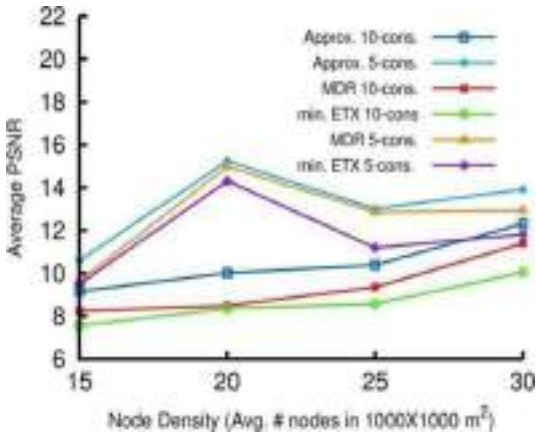


Fig. 3. Average PSNR for five and ten video connections

A subjective quality measure that tries to capture human impression relating to the video quality is the MOS (Mean Opinion Score). The result of the node density on the PSNR is shown in Fig. 3.We plot the average PSNR for five and ten synchronous video connections for various node densities and for video secret writing parameters of like with GOP size as five and Frame per second as thirty and Rate 273 kbps and Frame size as QCIF(176 X 144)pixels. We additionally plot the performance of our projected scheme (MDR). During this case, we tend to assume full knowledge of the topology so the state space where we tend to solve the optimum control problem is a superset of the state area once we collect the local estimates of ETX through the network.

4.2. Testbed Experiments

The experiment setup consists of an initial raw video processed using the H.264 encoder with a most GOP size of thirty frames. The traffic load ranges from two to twelve synchronous video flows, where the sender and receiver pairs are randomly selected. Each situation is repeated five times. To capture the effect of the ETX-based and MDR routing schemes on the user expertise, we tend to measure the average MOS because the number of synchronous video flows in the network will increase.

Fig. 4. shows that because the number of video connections within the network will increase, the average MOS for each schemes decreases. However, when the traffic load will increase, the MDR protocol computes multiple methods between the source and the destination nodes and is best in distributing the load across the network such the frames at the beginning of a GOP avoid congestion. Fig. 5.shows snapshots from video clip transmitted over the testbed below totally different traffic conditions for each the ETX and MDR protocol.

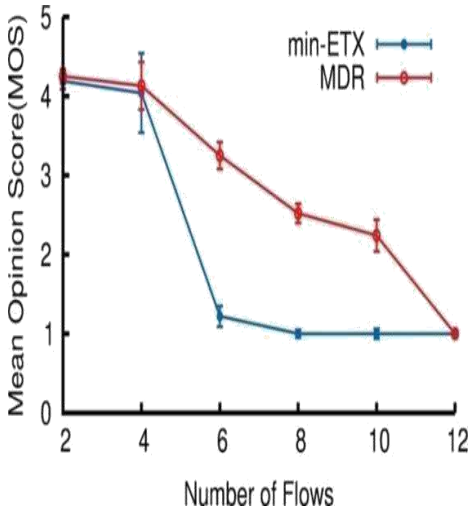


Fig. 4. Average value of MOS for a different number of concurrent video flows.

Fig. 5. User experience under different traffic loads.
(a)Video snapshot—MDR (two connections). (b) Video snapshot—ETX (two connections). (c) Video snapshot—MDR (eight connections). (d) Video snapshot—ETX (eight connections).

5. CONCLUSION

In this paper, a network that primarily carries video flows is taken into account. We try to understand the effect of routing the video flow on the end-to-end distortion. To account evolution of packet loss, an analytical model is built that ties video distortion to the underlying packet-loss probabilities. Using this model, the optimum route (in terms of distortion) is decided between a source and a destination node based on dynamic programming approach. The framework permits to formulate a routing policy for minimizing distortion, based on that we tend to design a practical routing theme that's evaluated via extensive simulations and testbed experiments. Our simulation study

shows that the distortion (in terms of PSNR) is decreased by 20 percent compared to ETX based routing. The future work of this paper is to minimize the users experienced degradation and to decrease the distortion more than 20%.

REFERENCES

[1] [1] T.Wiegand,G.J. Sullivan, G. Bjontegaard, and A. ,―Overview of the H.264/AVC video coding standard, vol. 13, no. 7, pp. 560– 576, Jul. 2003.

[2]

[3] [2] D.S.J.D.Couto,D.Aguayo,J.Bicket,andR.Morris, A high-throughput path metric for multi-hop wireless routing,‖ in Proc. 9 MobiCom, San Diego, CA, USA, Sep. 2003, pp. 134–146.

[4]

[5] [3] M. T. Ivrlac, L. U. Choi, E. Steinbach, and J. A. Nossek, ―Model and analysis of streaming video transmission over wireless fading channels,‖ Signal Process., Image Commun., vol. 24, no. 8, pp. 651–665, Sep. 2009.

[6]

[7] [4] S. Mao, Y. T. Hou, X. Cheng, H. D. Sherali, S. F. Midkiff, and Y.- Q. Zhang, ―On routing for multiple description video over wireless ad hoc networks, IEEE Trans. Multimedia, vol. 8, no. 5, pp. 1063– 1074, Oct. 2006.

[8]

[9] [5] S. Mao, X. Cheng, Y. T. Hou, and H. D. Sherali, Multiple description video multicast in wireless ad hoc networks, Mobile Netw. Appl., vol. 11, no. 1, pp. 63–73, Feb. 2006.

[10]

[11] [6] D. Kandris, M. Tsagkaropoulos, I. Politis, A. Tzes, and S.Kotsopoulos, Energy efficient and perceived QoS aware video routing over wireless multimedia sensor networks, AdHoc Netw., vol. 9, no. 4, pp. 591–607, 2011.

[12]

[13] [7] E. Kohler , R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek, The click modular router,‖ Trans. Comput. Syst., vol. 18, no. 3, pp. 263– 297, Aug. 2000.

[14]

[15] [8] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-Lossinduced distortion in decoded video," IEEE Trans. Circuits Syst Video Technol., vol. 16, no. 6, pp. 716–732, Jun. 2006.

[16]

[17] [9] D. Migliorini, E. Mingozzi, and C. Vallati, "Performance evaluating of H.264/SVC video streaming over mobile WiMAX," Comput. Netw., vol. 55, no. 15, pp. 3578–3591, Oct. 2011.

[18]

[19] [10] E. Alotaibi and B. Mukherjee, "A survey on routing algorithms for wireless ad-hoc and mesh networks," Comput. Netw.,vol.56,no.2,pp. 940–965, Feb. 2012.

[20]

[21] [12] L. Hanzo and R. Tafazolli, "A survey of QoS routing solution for Mobile ad hoc networks," IEEE Commun. Surveys Tuts., vol. 9, no.2, pp. 50–70, Apr. 2007.

[22]

[23] Proceeding Papers :

[24]

[25] [11] W. Wei and A. Zakhor, "Robust multipath source routing protocol (RMPSR) for video communication over wireless ad hoc networks," in Proc. IEEE ICME, Taipei, Taiwan, Jun. 2004, pp. 1379–1382.

[26]

[27] [12] J. Chen, S.-H. G. Chan, and V. O. Li, "Multipath routing for video Delivery over bandwidth-limited networks," IEEE J. Sel. Areas Commun., vol. 22, no. 10, pp. 1920–1932, Dec. 2010.

[28]

[29] [13] M. Garetto, T. Salonidis, and E. W. Knightly, "Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks," in Proc. IEEE IFOCOM,Barcelona,Spain,Apr.2006,pp 1–13.

[30]

[14] K. Medepalli and F. A. Tobagi, "Towards performance modeling of IEEE 802.11 based wireless networks: A unified framework and its applications," in Pro. IEEE INFOCOM, Barcelona, Spain, Apr. 2006

Three Tier Approach for Secure Data Transmission by Using Steganography, Logistic Maps with Genetic Algorithm and Visual Cryptography

Vamsi Krishna.Y , Varuni.R.V, Shivaranjini.A
Department of Information Science, Sri Krishna Institute of Technology, Bangalore-90
yvnvk.1229@gmail.com , varuni.venus@gmail.com, satheight@gmail.com

ABSTRACT-With advancement of technology, providing security to the data being transmitted is becoming a challenging task. So this paper mainly focuses on enhancing the existing steganography technique with additional feature of visual cryptography for the purpose of securing data hidden in image which is being accomplished through a 3 layered approach, first layer deals with high degree randomization of pixel selection for, LSB based steganography with logistic maps to reduce image degradation. The resulting stego-image undergoes shuffling based on genetic algorithm which builds up the second wall of security. Visual cryptography forms the third layer of security where transparencies of image are transmitted.

Keywords—Genetic Algorithm, Image Security, Logistic Maps, Steganography and Visual Cryptography.

1. INTRODUCTION

The method of concealing confidential messages within a media file that may be picture, vedio clip, audio is steganography. The purpose of steganography is to avoid illegal accessors from differentiating between Data Embedded Image and Plain Image.

Spatial domain, transform domain, spread spectrum and model based steganography are the types of image steaganography. Spatial domain deals with replacement of pixel bits in the image with the secret data. One of the spatial domain steganographic technique is Least Significant Bit(LSB) method. Least Significant Bit(LSB), as the name indicates, is method where LSB of selected pixels are embedded with bits of private data. In this case since their is less degree of randomized pixel selection the selected pixels have higher degree of possibilities of being selected from particular region in the image which results in notifiable degradation in image, as a solution logistic maps has been introduced whose main purpose to randomize pixel selection which results in reduction of degradation in image quality. Genetic algorithm will be implemented on the resulting stego-image to modify and shuffle the pixel location which will be discussed in detail as part of proposed methodology. The image with modified pixel location for the purpose of strengthening the security is broken into two shares before transmitting over network, this constitutes visual cryptography. This method has gained significant importance in the field of biometric security, watermarking, remote electronic voting, bank customer identification.

This paper is organized into four sections. Section I gives brief introduction to the concepts discussed in this paper. Section II deals with related work .Section III deals with discussion of problems in existing system that motivated to come up with this proposal. Section IV describes the various modules involved in the proposed system. Conclusion and future enhancement are discussed in section V.

2. RELATED WORK

Hamidreza et al. [1] proposes genetic algorithm based steganography approach to identify optimal location for embedding data in the image. [2] Mansi S Subhedar discusses the issues in steganography. Anandi et al. [3] proposes various visual cryptography schemes for secret images. Shirish kumbar et al. [4] proposes the use of visual cryptography in various fields like biometric, DNA etc. Rehana Begum R.D et al.[5] proposes the integration of LSB based steganography using genetic algorithm and visual cryptography for secured data hiding . Gokul et al [6]. Proposes a combination of visual cryptography and LSB encryption. Jeyamala chandrasekaran et al. [7] introduces the concept of logistic maps for optimal pixel selection. Fridrich et al. [9] proposes RS algorithm to detect LSB based data embedding in grayscale images. He also proposes F5 algorithm for steganalysis of data hidden in JPEG images et al. [10]. Sonaz Abdulla et al. [11] proposes a new visual cryptography technique of dividing images into transperencies before it is sent over network to improve security. Divya james et al. [12] proposes a visual cryptography based solution for phishing problem. Amritha et al. [13] proposes a genetic algorithm based steganography technique using discrete cosine transformation.

3. EXISTING SYSTEM

The current systems are not capable of higher degree randomized pixel selection in cover image for data embedding. This causes notifiable amount of image distortion which is easily prone to suspection, of data transmission through image. The proposed system provides solution to this problem by applying logistic maps which increases the randomization of pixel selection . In existing systems, after employing visual cryptography, at the decoder end on overlapping the shares, the original stego image is exposed as shown in Fig.2 thus resulting in reduction of security of stego image.

This problem is overcome by employing genetic algorithm which involves crossover and mutation of pixels on the stego image before visual cryptography is applied. As a result, at decoder end, on stacking the transparencies shuffled stego image is obtained and not the original stego image. Thus this enhances the security provided to the image concealing the private data.

4. PROPOSED METHODOLOGY

4.1 Logistic equation

Logistic equation portrays complex and chaotic behavior. A study on logistic maps provided a suitable equation for the purpose of key generation:

(1) $x(n) = 2 * (2 - (5 * x(n-1))) * x(n-1)$

Logistic equations are of great importance due to the fact that, very minute change in decimal part of input displays a huge variation in output. Without implementation of logistic equation, the image degradation is visible since pixels are not much randomly selected as shown in Fig.1. In proposed system, logistic equation (1) is incorporated to generate highly random values that are mapped with the pixel positions in the cover image, thus resulting in selection of pixels that are widely distributed and highly random in the cover image for the purpose of embedding data which overcomes the problem of image degradation as shown in Fig.2.


Original image


Pixel selection


Resulting image

Fig.1 Pixel selection without logistic maps.


Original image


Pixel selection


Resulting image

Fig.2 Randomized pixel selection with logistic maps.

4.2 Genetic algorithm

Existing systems embed data in the image and this image is given as input to visual cryptography as shown in Fig.3, which divides the image into two shares (explained in detail in the following module) sends it over network. If hacker obtains access to both the shares during transmission, on overlapping these shares, the hidden data can be extracted, the solution to this is genetic algorithm. Genetic algorithms are the class of algorithms that provide solutions for optimization and search problems. Crossover and mutation are the types under genetic algorithm that are made use of in the proposed system. Crossover can be further classified i.e., column shuffling and row shuffling. Column shuffling involves interchanging the pixels of two columns. Similarly row shuffling interchanges the pixel of two rows. In this paper row shuffling follows column shuffling. Mutation is a technique where inversion of pixel is achieved. Thus integration of these two techniques greatly enhances security provided to the stego-image. This method is incorporated in the proposed paper as shown in Fig.4.When hacker overlaps the two shares he obtains shuffled stego image and not the original stego image.


Share 1 Share2 Stego image

Fig.3 Before implementing genetic algorithm.


Share 1 Share 2 Shuffled stego image

Fig.4 After implementing genetic algorithm.

4.3Visual cryptography

The shuffled stego-image resulting from the previous module which is genetic algorithm if transmitted over

network any unauthorized individual who manages to access this image gets access to the whole image that contains the data this problem can be overcome by applying technique such as visual cryptography. Visual cryptography is a prominent technique that divides the image containing the message into two shares such that each of the share contains some part of data before transmitting over the network as shown in Fig.6. Thus provides security for the image containing the data. This method minimizes the risk of data being hacked by intruder. If the intruder manages to illegally access a share of the image during transmission, still the data cannot be revealed since it remains incomplete without the other share. This method is based on interpretation of pixels as binary bits. Each selected pixel is broken into eight subpixels, four pixels for each share. Once the shares are received at the receiver end, they are aligned by invoking reverse visual cryptography process to obtain the shuffled stego image.

4.4 Three Tier Approach for Secure Data Transmission by using Steganography, Logistic Maps with Genetic Algorithm and Visual Cryptography

The steps of the algorithm is as shown in Fig. 5

Sender Side Algorithm as shown in Fig.7

Input: Cover image, secret data
Output: Two shares of shuffled stego-image

Step1: Selecting random pixels through logistic maps.
Step2: Embedding secret data in previously selected pixels using LSB technique which is steganography.
Step3: Genetic algorithm is applied on resulting stego image which implements crossover and mutation.
Step4: Visual cryptography divides the shuffled stego image into two shares and sent through the network.

Receiver Side Algorithm as shown in Fig.8

Input: Shares of visual cryptography.
Output: Extraction of secret data.

Step1: Inverse visual cryptography method stacks the shares one over the other obtained at receiver end.
Step2: Output of the previous step is taken as input for inverse genetic algorithm to obtain original stego image.
Step3: Extraction of secret data from the stego image is achieved through inverse steganography process.

The overall design of the proposed system is depicted as shown below:

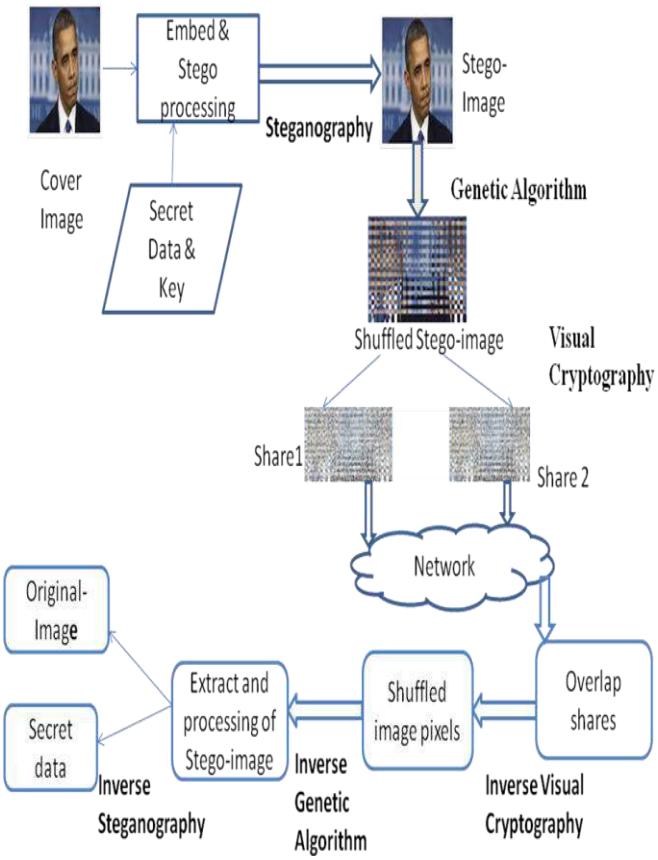


Fig.5 The Proposed Model

The following image depicts the working of simple visual cryptography without the application of genetic algorithm

Share 1 **Share 2** **Final**

Fig.6 Visual cryptography

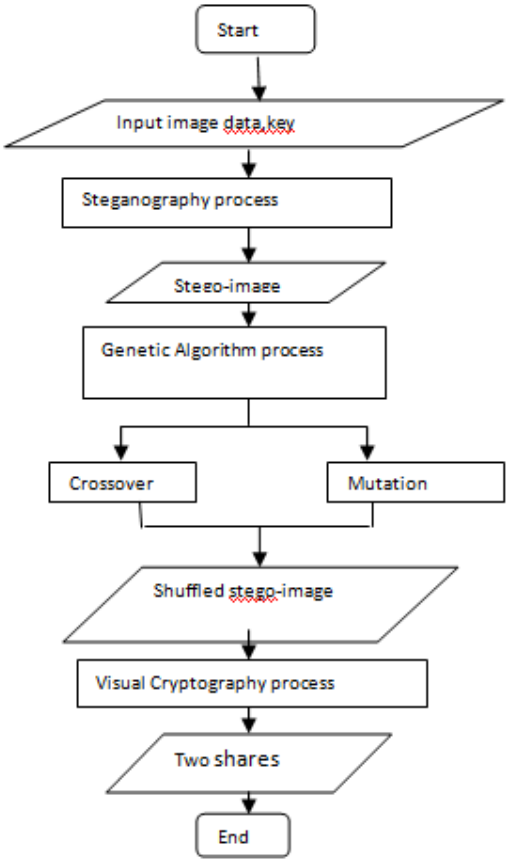


Fig.7 Sender side model

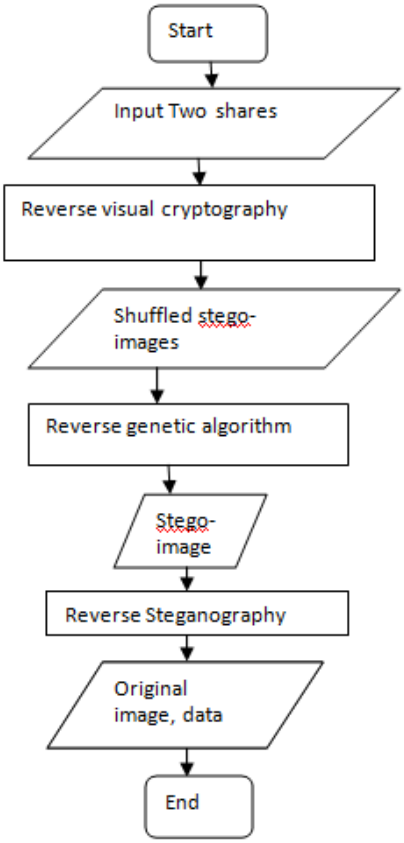


Fig.8 Receiver side model

5. CONCLUSION

This research focuses on enhancing the image security by adding a prominent feature which is visual cryptography to the existing system which integrates logistic maps for random pixel selection, LSB technique for data embedding, genetic algorithm to modify the pixel locations in stego-image.

Future work focuses on reduction of data loss due to imperfect alignment of transparencies and to minimize the interdependencies of shares during transmission.

REFERENCES

[1] Hamidreza Rashidy Kanan , Bahram Nazeri, "A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm", Expert Systems with Applications 41 (2014) 6123–6130.

[2] Mansi S Subhedar, Vijay.H.Mankar," Current Status and Key issues in Image Steganography", Computer Science Review,13-14, 2014, pp95-113.

[3] Anandi,S.Satthiyaraj," Embedded Visual Cryptography Schemes For Secret Images",(IJCSNS),vol.12,no.12,dec 2012.

[4] Miss.Shubhangi Rajanwar,Mr.Shirish Kumbar, Mr.Akshay Jadhav,Prof.Saba Siraj," Visual Cryptography For Image Privacy*",(IJCST)vol.3* issue2,Mar-Apr2015.

[5] Rehana Begum R.D,Sharaya Pradeep," Best Approach For LSB Based Steganography Using Genetic Algorithm And Visual Cryptography For Secure Data Hiding And Transmission Over Networks" *,(IJARCSSE), vol.4* ,issue6, June2014.

[6] Gokul.M, Umesh Babu.R, Shriram K Vasudevan, Deepak Karthik," Hybrid Steganography Using Visual Cryptography And LSB Encryption Method", *(IJCA),[0975-8887],vol59*,no.14,Dec2012

[7] Jeyamala Chandrasekaran , Girija Arumugam, Deepthi Rajkumar, "Ensemble Of Logistic maps With Genetic Algorithm For Optimal Pixel Selection In Image Steganography*",(ICECS),2015*.

[8] https://www.irjet.net/archives/V2/i5/IRJET-V2I5224.pdf

[9] Fridrich J .,Goljan M and Du R,"Reliable Detection Of LSB Steganography In Color And Grayscale Images ", Proceedings of Workshop on Multimedia and Security, Ottawa,pp.27-30, october 5 2001.

[10] J. Fridrich,M.Goljan and D.Hogea,"Steganalysis Of JPEG Images:Breaking The F5 Algorithm",In proc. Of the ACM Workshop on Multimedia and Security,2002.Sozan Abdulla, "New Visual Cryptography Algorithm For Colored Image*", Journal of computing, volume 2*,issue 4, April 2010 .

[11] Divya James, Mintu Philip, "A Novel Anti Phishing framework based on Visual Cryptography", IEEE, 2011.

[12] Amritha Khamrui,J.K mandal," A Genetic Algorithm Based Steganography Using Discrete Cosine Transformation (GASDCT),*International Conference on Computational Intelligence: Modeling, Techniques and Applications(CIMTA)*,2013,pp.105-111.

Author Profile

Y Vamsi Krishna is working as an Associate Professor in the Department of Information Science and Engineering. Has completed his Mtech from VTU with 10 years of experience in teaching .His area of interest is mainly in image processing and artificial intelligence for national and international conference.

Varuni R.V, born on 9th july,1994 pursuing B.E in information science. Studying at Sri Krishna Institute of Technology, Chikkabanavara, Karnataka, India. Areas of Interest are Robotics and Artificial Intelligence

Shivaranjini A, born on 15th october,1994 pursuing B.E in information science. Studying at Sri Krishna Institute of Technology, Chikkabanavara, Karnataka, India. Areas of Interest are Nano Technology and Artificial Intelligence

Greedy Forwarding for Wireless Sensor Networks with Guaranteed Delivery

Namita Mirjankar, Siddhitha S Pai, Lohit Patil
BACHELOR OF COMPUTER APPLICATION, KLES's BCA, hubballi-580031
namitamir@gmail.com, Sidhitapai@gmail.com, princelohit95@gmail.com

ABSTRACT : The topology of a remote sensor system changes as a few sensors come up short on force, fall flat or Join the system; which might bring about loss of information or the velocity of exchange of information backs off. This issue is solved using Greedy Algorithm and the Hop Count Reduction (HCR) plan is used as a short-listening so as to slice procedure to decrease the steering bounces to the neighbor's movement, while the intersection navigation (IN) component is proposed to acquire the best moving bearing for limit traversal with the selection of most brief way criterion. The verification of accuracy for the GAR plan is additionally given in this paper.

Keywords – Greedy Routing Algorithms, Hop Count Reduction (HCR), localized algorithm, unit disk graph, void problem, Wireless Sensor Networks.

I. INTRODUCTION

A Wireless sensor network system (WSN) comprises of sensor hubs (SNs) with remote correspondence abilities for particular detecting undertakings. Because of the constrained accessible assets, proficient configuration of confined multi-jump steering conventions turns into a urgent subject inside of the WSNs. The most effective method to ensure conveyance of bundles is viewed as an imperative issue for the restricted steering calculations. The surely understood Greedy Forwarding (GF) calculation is viewed as a prevalent plan with its low directing overheads. Be that as it may, the void issue, which makes the GF strategy not able to locate its next closer hop to the destination, will bring about the GF calculation neglecting to ensure the conveyance of information parcels. A few steering calculations are proposed to either resolve or diminish the void issue, which can be ordered into non-chart based and diagram based plans. In the non-diagram based calculations the natural plans as proposed in build a two-bounce neighbor table for executing the GF calculation. The system flooding component is received with in the GRA and PSR plans while the void issue happens. There likewise exist steering conventions that receive the backtracking technique at the event of the system gaps, (for example, GEDIR, DFS and SPEED). The directing plans as proposed by ARP and LFR retain the steering way after the void issue happens. Besides, other directing conventions, (for example, PAGER, NEAR, DUA, INF, and YAGR) proliferate and redesign the data of the watched void hub with a specific end goal to decrease the likelihood of experiencing the void issue. By abusing these directing calculations, in any case, the void issue must be either halfway mitigated or determined with impressive steering overheads and critical focalizing time.

Then again, there are examination chips away at the configuration of diagram based steering calculations to manage the void issue. A few steering plans as overviewed in receive the planar diagram got from the unit plate chart (UDG) as their system topologies, for example, GPSR , GFG , Compass Routing II , AFR [, GOAFR GOAFR+, GOAFR++, and GPVFR . For directing the above planar diagram based calculations, the planarization system is required to change the hidden system chart into the planar chart. The Gabriel diagram (GG) and the relative neighborhood chart (RNG) are the two usually utilized confined planarization strategies that relinquish some correspondence joins from the UDG for accomplishing the planar diagram. By the by, the utilization of the GG and RNG diagrams has noteworthy pitfalls because of the evacuation of basic correspondence joins, prompting longer directing ways to the destination..

II. IMPORTANAT TERMINOLOGIES

A. Wireless Sensor Network

A wireless sensor network (WSN) (in some cases called a wireless sensor and actuator network(WSAN)) are spatially disseminated self-governing sensors to screen physical or ecological conditions, for example, temperature, sound, weight, and so forth and to agreeably go their information through the system to a principle area. The more present day systems are bi-directional, likewise empowering control of sensor action. The improvement of remote sensor systems was propelled by military applications, for example, front line observation; today such systems are utilized as a part of numerous modern and buyer applications, for example, mechanical procedure checking and control, machine wellbeing observing, et cetera.

The WSN is worked of "hubs" – from a couple to a few hundreds or even thousands, where every hub is associated with one (or some of the time a few) sensors. Each such sensor system hub has regularly a few sections: a radio handset with an inward receiving wire or association with an outside reception apparatus, a microcontroller, an electronic circuit for interfacing with the sensors and a vitality source, as a rule a battery or an installed type of vitality collecting. A sensor hub may differ in size from that of a shoebox down to the extent of a grain of dust, albeit working "bits" of honest to goodness infinitesimal measurements have yet to be made. The expense of sensor hubs is comparably variable, running from a couple to several dollars, contingent upon the many-sided quality of the individual sensor hubs.

Size and cost imperatives on sensor hubs result in comparing requirements on assets, for example, vitality, memory, computational velocity and interchanges data transfer capacity. The topology of the WSNs can fluctuate from a straightforward star system to a progressed multi-jump remote cross section system. The proliferation system between the bounces of the system can be directing or flooding.

B. Greedy Routing Algorithms:

Covetous Algorithm is a numerical procedure that searches for a basic and simple to actualize answer for complex, multi-step issues by choosing which next step will give the most evident advantage. Such a calculation is called ravenous in light of the fact that while the ideal answer for each littler case will give a quick yield, the calculation does not consider the expansive issue as entirety. Once a choice has been made it is never reexamined.

In this paper, a conveyance ensured area free steering convention, termed LF-GFG, is proposed for a remote sensor system with evolving topology. We first depict the system multivalued installing convention to guide every hub and every connection in the system to numerous virtual hubs and various virtual connections, individually, to constitute a virtual system in a plane and exhibit the virtual system planarization convention to get the associated spreading over planar sub-diagram of the virtual system. At that point, LF-GFG advances a bundle utilizing the covetous face-avaricious (GFG) calculation in light of the virtual system and the associated crossing planar sub-diagram. As the system topology changes, the upkeep plan reproduces an associated traversing planar sub-diagram of the virtual system, utilizing simply nearby data, just if the spreading over planar sub-chart gets to be separated. In this manner, dissimilar to existing area free directing conventions, LF-GFG requests just lightweight upkeep costs as the system topology changes because of hub expansion or evacuation. Recreations in the system test system NS-2 demonstrate that LF-GFG has great execution regarding the development message overhead, the support time and message overhead, and the bundle conveyance rate while guaranteeing moderate directing dormancy costs.

C. Hop Count Reduction(HCR):

In PC organizing, a hop is one bit of the way in the middle of source and destination. Information bundles go through extensions, switches and entryways in transit. Every time bundles are gone to the following gadget, a hop happens. Since store and forward and different latencies are caused through every jump, an expansive number of hop in the middle of source and destination suggests bring down continuous execution.

The hop check alludes to the quantity of middle of the road gadgets (like switches) through which information must go in the middle of source and destination, as opposed to streaming straightforwardly over a solitary wire. Every switch along the information way constitutes a bounce, as the information is moved starting with one Layer 3 arrange then onto the next. Hop tally is accordingly an essential estimation of separation in a system. Hop check is a harsh measure of separation between two hosts. A jump tally of n implies that n passages isolate the source host from the destination host. Without anyone else, this metric is, be that as it may, not valuable for deciding the ideal system way, as it doesn't contemplate the rate, burden, unwavering quality, or idleness of a specific bounce, however just the aggregate tally. Every time an able gadget gets these parcels, that gadget changes the bundle, augmenting the Hop tally by one. What's more, the gadget thinks about the bounce mean something negative for a period as far as possible and tosses the bundle if its jump number is too high. This keeps bundles from interminably bobbing around the system in the occasion of steering blunders. Switches are equipped for overseeing jump tallies, however different sorts of middle of the road gadgets are most certainly not.

III. MODULES USED IN BUILDING THE SYSTEM

A. NETWORKING MODULE:

Client server figuring or systems administration is a dispersed application engineering that segments undertakings or workloads between administration suppliers (servers) and administration requesters, called clients. Frequently clients and servers work over a PC system on independent equipment. A server machine is a superior host that is running one or more server projects which impart its assets to customers. A customer additionally shares any of its assets; Clients along these lines start correspondence sessions with servers which anticipate (listen to) approaching solicitations.

B. BOUNDRY EVOLUTION MODULE:

The RUT plan is received to take care of the limit discovering issue, and the mix of the GF and the RUT plan (i.e., the GAR convention) can resolve the void issue, prompting the ensured bundle conveyance. The meaning of limit and the issue articulation are portrayed as takes after: Definition 1 (limit). On the off chance that there exists a set B such that 1) the hubs in B frame a basic unidirectional ring and 2) the hubs situated on and inside the ring are detached with those outside of the ring, B is meant as the limit set and the unidirectional ring is known as a limit.

C. Greedy Anti-void Traversal module.

The goal of the GAR convention is to determine the void issue such that the bundle conveyance from NS to ND can be ensured. Before plunging into the point of interest definition of the proposed GAR calculation, a basic illustration is depicted so as to encourage the comprehension of the GAR convention, the information parcels started from the source hub NS to the destination hub ND will land in NV in light of the GF calculation. The void issue happens as NV gets the parcels, which prompts the reception of the RUT plan as the sending procedure of the GAR convention. A circle is shaped by focusing at SV with its sweep being equivalent to half of the transmission range R/2.

D. Partial UDG Construction (PUC) Mechanism

The PUC component is focused to recuperate the UDG linkage of the limit hub Ni inside of a non-UDG system. The limit hubs inside of the proposed GAR convention are characterized as the SNs that are used to handle the parcel conveyance in the wake of experiencing the void issue .Therefore, leading the PUC component just by the limit

hubs can moderate system assets than most. The PUC instrument of the current flooding-based plans requires data from all the system hubs.

E. The performance Evolution module

The execution of the proposed GAR calculation is assessed and contrasted and other existing confined plans by means of recreations, including the reference GF calculation, the planar chart based GPSR and GOAFR++ plans, and the UDG-based BOUNDHOLE calculation. It is noticed that the GPSR and GOAFR++ plans that receive the GG planarization method to planarize the system diagram are spoken to as the GPSR (GG) and GOAFR++ (GG) calculations, while the variations of these two plans with the CLDP planarization calculation are signified as the GPSR (CLDP) and GOAFR++ (CLDP) conventions.designation.

IV. PROPOSED GREEDY ANTI-VOID ROUTING (GAR) PROTOCOL

The target of the GAR convention is to determine the void issue such that the bundle conveyance from NS to ND can be ensured..

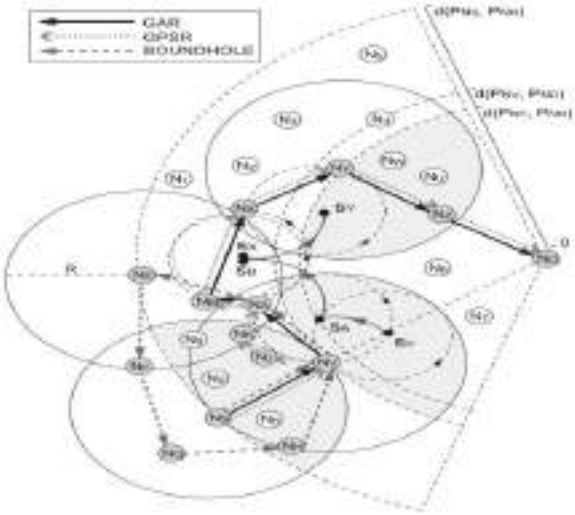


Fig.1. Example routing paths constructed by using the GAR, the GPSR, and the BOUNDHOLE algorithms under the existence of the void problem

Before plunging into the subtle element detailing of the proposed GAR calculation, a starting sample is depicted keeping in mind the end goal to encourage the comprehension of the GAR convention.

As appeared in Fig. 1, the information parcels started from the source hub NS to the destination hub ND will touch base in NV in light of the GF calculation. The void issue happens as NV gets the parcels, which prompts the reception of the RUT plan as the sending methodology of the GAR convention. A circle is shaped by focusing at sV with its span being equivalent to half of the transmission range R=2. The circle is pivoted at NV and begins to lead counterclockwise moving until a SN has been experienced by the limit of the circle, i.e., NA, as in Fig. 1. Therefore, the information bundles in NV will be sent to the experienced hub NA. In this way, another equivalent estimated circle will be shaped, which is focused at sA and pivoted at hub NA. The counterclockwise moving methodology will be continued so as to choose the following jump hub, i.e., NB for this situation. Likewise, same procedure will be performed by other halfway hubs, (for example, NB and NX) until the hub NY is achieved, which is considered to have a littler separation to ND than that of NV to ND. The traditional GF plan will be continued at NY for conveying information parcels to the destination hub ND. As an outcome, the subsequent way by embracing the GAR convention gets to be fNS;NV ;NA; NB;NX;NY ;NZ;NDg.

A. Proposed Rolling-Ball UDG Boundary Traversal (RUT) Scheme

The RUT plan is embraced to take care of the limit discovering issue, and the blend of the GF and the RUT\ plan (i.e., the GAR convention) can resolve the void issue, prompting the ensured parcel conveyance. The meaning of limit and the issue explanation are depicted as takes after:

Definition 1 (limit):

On the off chance that there exists a set B _ N such that

1) The hubs in B shape a basic unidirectional ring and

2) the hubs situated on and inside the ring are disengaged with those outside of the ring, B is meant as the limit set and the unidirectional ring is known as a limit.

i. Initialization Phase :

No calculation can be executed without the calculation particular trigger occasion. The trigger occasion inside of the RUT plan is known as the beginning stage (SP). The RUT plan can be instated from any SP, which is characterized as takes after:

Definition 2 (moving ball):

Given Ni 2 N, a moving ball RBNiðsi;R=2Þ is characterized by 1) a moving circle pivoted at PNi with its middle point at si 2 IR2 and the span equivalent to R=2, and 2) there does not exist anyNk 2 Nlocated inside the moving ball as fRB_Niðsi;R=2Þ \ Ng ¼ ;, where RB_Niðsi;R=2Þ indicates the open plate inside of the moving ball.

Definition 3(beginning stage):

The SP of Ni inside of the RUT plan is characterized as the middle point si 2 IR2 of RBNiðsi;R=2Þ. As appeared in Fig. 2, every hub Ni can check if there exists a SP since the moving ball RBNiðsi;R=2Þ is limited by the transmission scope of Ni.

As indicated by Definition 3, the SPs ought to be situated on the circle focused at PNi with a span of R=2. As will be demonstrated in Lemmas 1 and 2, all the SPs will bring about the red strong blossom molded curves, as in Fig. 2. It is seen that there ought to dependably exist a SP, while the void issue happens inside of the system, which will be clarified in Section 3.2. At this underlying stage, the area si can be chosen as the SP for the RUT plan.

ii. Boundary Traversal Phase

Given si as the SP connected with its RBNiðsi;R=2Þ pivoted at Ni, either the counterclockwise or clockwise moving bearing can be used. As appeared in Fig. 2, RBNiðsi;R=2Þ is moved counterclockwise until the following SN is come to (i.e., Nj in Fig. 2). The unidirectional edge Eij ¼ ðPNi; PNj Þ can along these lines be developed. Another SP and the correspondingrolling ball pivoted at Nj (i.e., sj and RBNjðsj;R=2Þ) will be doled out, and thusly, the same technique can be directed constantly. 3.1.3 Termination Phase The end condition for the RUT plan happens while the primary unidirectional edge is returned to. As appeared in

Fig. 2, the RUT plan will be ended if the edge Eij is gone to again after the edges Eij, Ejk, Ekl, Elm, and Emi are crossed. The limit set started from Ni can in this manner be gotten as B ¼ fNi;Nj;Nk;Nl;Nmg.

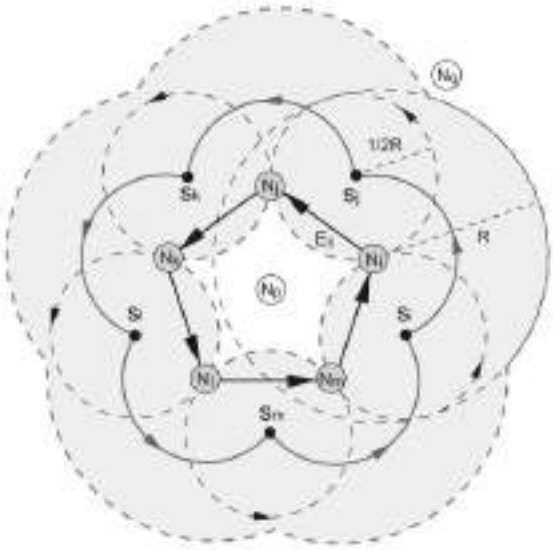


Fig2. The proposed RUT scheme.

B. Detail Description of Proposed GAR Protocol

As appeared in Fig. 1, the bundles are expected to be conveyed from NS to ND. NS will choose NV as the adopting so as to follow bounce hub the GF calculation. In any case, the void issue denies NV to keep using the same GF calculation for bundle sending. The RUT plan is thusly utilized by relegating a SP (i.e., sV) connected with the moving ball RBNV ðsV ;R=2þ pivoted at NV . As showed in Fig. 1, sV can be situated on the associating line in the middle of NV and ND with R=2 far from NV . It is seen that there dependably exists a SP for the void hub ðNV Þ since there shouldn't have any SN situated inside of the blue-shaded district (as in Fig. 1), which is sufficiently extensive to fulfill the necessities, as in Definitions 2 and 3. The RUT plan is used until NY is come to (in the wake of crossing NA, NB, and NX). Since dðPNY ; PND Þ <dðPNV ; PND Þ, the GF calculation is continued at NY , and the following bounce hub will be chosen as NZ. The course from NS to ND can thusly be developed for parcel conveyance. In addition, if there does not exist a hub NY such that dðPNY ; PND Þ <dðPNV ; PND Þ inside of the limit traversal stage, the RUT plan will be ended in the wake of returning to the edge EVA. The outcome demonstrates that there does not exist a directing way in the middle of NS and ND.

C. Proof of Correctness

In this segment, the accuracy of the RUT plan is demonstrated keeping in mind the end goal to take care of Problem 2, while the GAR convention is likewise demonstrated for determining the void issue (i.e., Problem 1) with a specific end goal to ensure parcel conveyance. Actuality1. A straightforward shut bend is shaped by navigating a point on the fringe of a shut filled 2D geometry with altered introduction.

Lemma1.

All the SPs inside of the RUT plan frame the fringe of a shape that outcomes from covering the shut circles DðPNi;R=2Þ for all Ni 2 N, and the other way around.

Confirmation:

Based on Definitions 2 and 3, the arrangement of SPs can be acquired as S¼R1 \ R2¼fsi jksi_PNik¼R=2; 9Ni2N; si 2 IR2g \ fsj j ksj _ PNjk _ R=2; 8 Nj 2 N; sj 2 IR2g by receiving the 1) and 2) rules inside of Definition 2. Then again, the outskirt of the subsequent shape from the covered shut plates DðPNi;R=2Þ for all Ni 2 N can be signified as _ ¼ Q1 _ Q2 ¼ SNi2N CðPNi;R=2Þ _ SNi2N DðPNi;R=2Þ, where CðPNi;R=2Þ and DðPNi;R=2Þ speak to the circle and the open circle focused at PNi with a range of R=2, separately. It is clear to notice that R1 ¼ Q1 and R2 ¼ Q0 2, which bring about S ¼ _. It finishes the verification. tu

Lemma2.

A straightforward shut bend is framed by the direction of the SPs.

Confirmation:

Based on Lemma 1, the direction of the SPs formsthe fringe of the covered shut circles DðPNi;R=2Þ for all Ni 2 N. Additionally, the outskirt of a shut filled 2D geometry is a basic shut bend as indicated by Fact 1. In this way, a basic shut bend is built by the direction of the SPs, e.g., the strong bloom formed shut bend, as in Fig. 2. It finishes the verification.

Hypothesis1.

The limit discovering issue (Problem 2) is determined by the RUT plan.

Confirmation:

Based onLemma2, the RUT scheme can draw a basic shut bend by turning the moving balls RBNiðsi;R=2Þ pivoted at PNi for all Ni 2 N. The shut bend can be separated into curve fragments Sðsi; sjþ, where si is the beginning SP connected with Ni, and sj is the grapple point while turning the RBNiðsi;R=2Þ pivoted at PNi . The curve fragments Sðsi; sjþ can be mapped into the unidirectional

Fig. 2. The proposed RUT plan.edgesEij ¼ ðPNi; PNj Þ for all Ni, Nj 2 U, where U _ N. Because of the coordinated mapping between Sðsi; sjþ and Eij, a straightforward unidirectional ring is developed by Eij for all Ni, Nj 2 U.

As per the RUT plan, there does not exist any Ni 2 N inside of the zone crossed by the moving balls, i.e., inside the light blue area, as in Fig. 2. For all Np 2 N situated inside the straightforward unidirectional ring, the littlest separation from Np to Nq, which is situated outside of the ring, is more prominent than the SN's transmission range R. Hence, there does not exist any Np 2 N inside the

straightforward unidirectional ring that can correspond with Nq 2 N situated outside of the ring.

In view of Definition 1, the set U is indistinguishable to the limit set, i.e., U ¼ B. It finishes the evidence.

Hypothesis 2.

The void issue (Problem 1) in UDGs is understood by the GAR convention with ensured bundle conveyance.

Evidence.

With the presence of the void issue happened at the void hub NV , the RUT plan is used by starting a SP ðsV Þ with the moving ball RBNV ðsV ;R=2þ pivoted at NV . The RUT plan inside of the GAR convention will lead limit (i.e., the set B) traversal under the condition that dðPNi; PNDÞ _ dðPNV ; PND Þ for all Ni 2 B. On the off chance that the limit inside of the hidden system is totally voyage taking into account Theorem 1, it shows that the SNs inside the limit (e.g., NV) are not fit for speaking with those situated outside of the limit (e.g., ND). The outcome demonstrates that there does not exist a course from the void hub ðNV Þ to the destination hub ðNDÞ, i.e., the presence of system segment. Then again, if there exists a hub NY such that dðPNY ; PND Þ <dðPNV ;

Application:

Because of the low energy consumption and less amount of packet loss greed routing algorithm is used for the transfer of packets. Main advantage of greedy routing algorithm is effective routing when compare to other algorithms.

V. Conclusion

A In this paper, a UDG-based GAR convention is proposed to determine the void issue brought about by the customary GF calculation. The RUT plan is embraced inside of the GAR convention to take care of the limit discovering issue, which brings about ensured conveyance of information bundles under the UDG systems. The BM and the IMS are additionally proposed to vanquish the computational issue of the moving component in the RUT plan, shaping the immediate mappings between the information/yield hubs. The proposed GAR calculations can promise the conveyance of information bundles under the UDG system.

References

[31] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor networks," Proc. ACM MobiCom, pp. 263-270, Aug. 1999.

[32] G.G. Finn, "Routing and Addressing Problems in Large Metropolitan-Scale Internetworks," Technical Report ISI/RR-87-180, Information Sciences Inst., Mar. 1987.

[33] B. Karp and H.T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," Proc. ACM MobiCom, pp. 243-254, Aug. 2000.

[34] Hannes Frey and Ivan Stojmenovic "On Delivery Guarantees of Face and Combined Greedy-Face Routing in Ad Hoc and Sensor Networks" 2006.

[35] Wen-Jiunn Liu, Student Member, IEEE, and Kai-Ten Feng, Member, IEEE for " Greedy Routing with Anti-Void Traversal for Wireless Sensor Networks" 2009

[36] YanSun *University of Kentucky*, for "EFFICIENT GREEDY-FACE-GREEDYGEOGRAPHIC ROUTING PROTOCOLS INMOBILE AD HOC AND SENSORNETWORKS " 2012.

[37] Yuan-Po Cheng, Yao-Jen Tang, and Ming-Jer Tsai for" LF-GFG: Location-Free Greedy-Face-Greedy Routing with Guaranteed Delivery and Lightweight Maintenance Cost in a Wireless Sensor Network with Changing Topology"2013.

Biographies and Photographs

Siddhitha S Pai is perceiving the BCA degree from KLES'S BCA Hubballi, P.C.Jabin Science college campus, Hubballi.

Lohit patil is perceiving the BCA degree from KLES'S BCA Hubballi, P.C.Jabin Science college campus, Hubballi.

Smt.Namita Mirjankar is working as a professor at KLES'S BCA Hubballi, P.C.Jabin Science college campus, Hubballi.

A Survey On SOA Governance Scope, Objectives And Its Policies

Pushpalatha B, Prof R Jaya
PG Student, Sr. Assistant Professor
Department of Computer Science Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India
pushpa41993@gmail.com, jayamanojkumar@gmail.com

ABSTRACT: Every enterprise that has been using IT as a business enabler is thinking of, discussing about or experimenting with Service Oriented Architecture. However, a key issue that has been identified in the SOA context is the governance of the Service Oriented Architectures. Service-oriented architecture (SOA) governance transcends simply designing and developing Web services and implementing SOA registries, repositories and other like tools. Robust organizational groundwork is required for an SOA implementation to succeed. Every enterprise that has some relation to Information Technology (IT), be it as consumer or as provider, in Service Oriented Architecture (SOA). the paper attempts to give the structure, roles, processes, features, scope and objectives that will together help set up an model to see the best SOA governance which enable --agility, efficiency and adaptability.

Keywords - Service-oriented architecture, IT governance, SOA governance, SOA governance model, Governance policies, Life cycle Governance.

I. INTRODUCTION

The Prior to looking into IT Governance and SOA Governance, let's try and understand the need for Governance by briefly analyzing the SOA paradigm and how it is different from traditional software application development. SOA is an architecture style and a way of organizing the solution to foster reuse, growth and interoperability. Business-aligned IT services form the first-order construct in a Service Oriented Enterprise in contrast to applications in a traditional IT world. These services are then orchestrated and assembled to support Enterprise Business Processes.

1.1 WHAT IS SOA GOVERNANCE?

The term SOA refers to Service Oriented Architecture and SOA Governance is related to governance of a Service-Oriented Enterprise that uses a business-driven, services based approach to IT solution design, development, operation and management.

1.2 KEY SOA GOVERNANCE POLICIES

Strategic governance: These would include anything related to SOA strategy, including who is responsible for creating it, changing it and implementing it.

Enterprise governance: What architectural standards should the enterprise follow? Who defines them? What is the process for adding new standards or products into the approved list? When shared service?

Program governance: These policies are used to define and approve business services, prioritization of service development and deployment, service ownership, and which information systems are the "systems of record" for each particular piece of information.

Release management: When should services and SOA infrastructure be released into production use? When should services be retired? What are the acceptance criteria for releasing the services for use? These are part of release management governance policies.

Service life cycle management: These are policies that determine when and how the services are identified, how service contracts are developed, reviewed and approved, and how services are developed and put into production.

Change management: Changes to services and SOA infrastructure must be controlled in order to maintain a "clean" and well-organized SOA environment.

II THE DOMAINS OF SOA GOVERNANCE

SOA governance can be classified into multiple categories. These include the following:

Architecture governance: Architecture governance is about enforcing architecture principles, leading practices and standards so that architectural decisions and policies are followed throughout the enterprise in a consistent manner.

Service life cycle management governance: Organizations typically find themselves possessing dozens—if not hundreds—of services created by many different workgroups and departments.

Policy life cycle management governance: Governance is about defining policies and enforcing these policies. But the policies themselves have their own life cycles.

2.1 SOA GOVERNANCE – THE SCOPE

SOA Governance must govern the entire service cycle from strategy to design to development to operations and subsequent SLA management. The governance

considerations across the entire SOA lifecycle are represented below, fig 2.1

Let's now look at what a Governance model.[4]

Fig2.1: governance model

SOA Investment Prioritization – The governance model must specify how SOA business cases are put forward and who has the final authority in prioritizing and approving SOA projects.

SOA Funding – Deciding upon this aspect within the governance model helps understand who must fund a particular project/investment and also lays down chargeback policies for service usage.

SOA Infrastructure Management – The governance model must specify the entity who will decide on the hardware and the software components required for building the services platform and the process for building the SOA infrastructure over time.

Service Ownership – This function establishes ownership for business services in the Enterprise Service Mosaic. The Business Service Owner is responsible for upgrades and enhancement of services and in ensuring that the service meet with the Quality of Service (QoS) requirements specified in its contract.

SOA Risk Management – This function helps to define the processes for continuous monitoring of risks in a SOA program in order to take immediate corrective action.

SOA Business Value – The function encompasses the metrics for measuring business value and systematic process for capturing metrics, quantifying and publishing business value from SOA are defined.

SOA Competency Management – The different roles required for a SOA program and the corresponding skills must be defined. This is then used to drive SOA training initiatives and to build the required competencies and skill base within the organization.

SOA Architecture and Principles – This function aims at defining/refining the SOA Reference Architecture, Principles and Standards that the service design and development uses.

Service Identification – Identification of services with right level of granularity is the most critical aspect of SOA. The services must be coarse-grained and defined keeping reuse and usage in different business contexts in mind.

SOA Change Management – The governance model must specify the process to make changes to services in production and on maintaining multiple versions of a service.

Service Registry and Taxonomy– Specifies who is responsible for publishing services into the registry and for defining service taxonomy to enable one to easily search for services.

SOA Practices and Procedures – Best Practices with respect to building business services, integration services, technical services and use of pattern-based approaches are defined. Processes are defined to ensure that learning from every project goes into the SOA knowledge base and it is institutionalized by rolling into SOA Practices and Procedures.

SOA Reuse – This function helps define structures and process to ensure that services are reused wherever available and puts in place the necessary infrastructure to facilitate reuse in an organization.

SOA (Build v/s Buy) – Defines the process for making Build v/s Buy decision.

Service Monitoring and SLA Management – This function defines the aspects of a service that must be controlled and monitored. Process for reporting and managing SLA exceptions is defined.

SOA Policies and Security – It is critical to control access to services and protect the data that is exchanged by enforcing appropriate policies. Security policies must be tightly controlled and the process and organization structure to support it are defined.

III OBJECTIVES OF SOA GOVERNANCE

Flexible, Business-aligned IT Enterprise

Business Processes and applications are decomposed into modular components in a Service Oriented Enterprise. As we start rethinking components for business and applications, we must, over time, build a portfolio of process, business and application services that is in tune with the business strategy, business architecture and mission of the enterprise.

SOA Platform Realization

In order to accrue the benefits of SOA, the right set of SOA platform infrastructure elements need to be selected. The SOA platform implementation can yield cost benefits, flexibility and architecture simplification; however, tightly controlled governance is required to implement it in practice.

Right Grained Business Service Identification And Design

Business excellence and continuous process improvement can be achieved through the re-use and continuous improvement of services. The SOA design must ensure that changes to business processes and services, leading to process improvements, can be implemented easily by localizing the change to a set of services.

Implementation Standardization

Like the obvious objective of any governance system, enforcing the standard way of implementation by way of usage of patterns, best-practices and pre-built templates should be aimed at in order to control cost and complexity and to increase reliability and quality of the end solution.

IV BASICS OF GOOD GOVERNANCE

Accountability - government is able and willing to show the extent to which its actions and decisions are consistent with clearly-defined and agreed-upon objectives. Accountability can be both an end in itself – representing values – and a means towards development of efficient and effective organizations. Accountability is a key way to ensure that resources and decision making powers are used appropriately.

Efficiency and Effectiveness - government strives to produce quality public outputs, including services delivered to citizens, at the best cost, and ensures that output meets the original intentions of policymakers.

Transparency - government actions, decisions and decision-making processes are open to an appropriate level of scrutiny by other parts of government, civil society and, in some instances, outside institutions and governments.

Responsiveness - government has the capacity and flexibility to respond rapidly to societal changes. It takes into account the expectations of civil society in identifying the general public interest, and is willing to critically re-examine the role of government.

Forward vision - government is able to anticipate future problems and issues based on current data and trends and can develop policies that take into account future costs and anticipated changes.

Rule of law - refers to the institutional process of setting, interpreting and implementing laws and other regulations. It means that decisions taken by government must be founded in law.

Participation - of governed subjects is key yardstick that is used to measure the overall success of the governance system. Participation gives government access to important information about the needs and priorities of individuals, communities and private businesses.

To implement a structure of governance with three constituent roles of Legislative, Executive and Judiciary with proper checks and balances fig 4

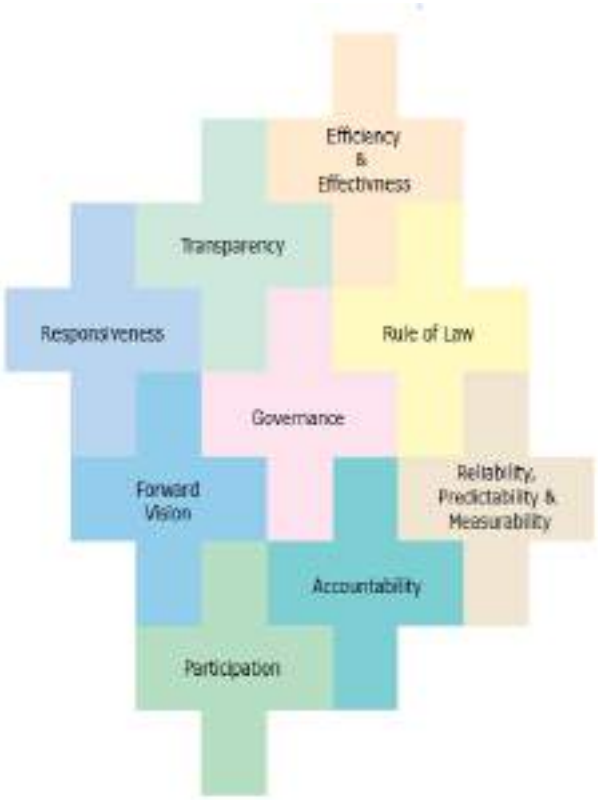


Fig 4: Basic Features of Good Governance

V CONCLUSION

By implementing SOA in an evolutionary manner through incremental development and deployment of business applications and reuse of business components, organizations have a much better chance of building the right architecture and adopting the right practices needed to bring a successful SOA vision to fruition.

.

REFERENCES

[1]Luthria, H., Rabhi, F.: Service-oriented computing in practice – an agenda for research into the factors influencing the organizational adoption of service oriented architectures. J. Theor. Appl. Electron. Commer. Res. 4(1), 39–56 (2009)

[2] Beimborn, D., et al.: The role of IT/business alignment for achieving SOA business value - proposing a research model. In: Americas Conference on Information Systems (AMCIS). AIS Electronic Library (AISeL)(2009)

[3]www.findwhitepapers.com/Technology/Data Management

[4] www.oracle.com/us/.../soa/soa-lifecycle-governance-wp-

971496.pdf
[5]Shankar Kambhampaly, *"Service–Oriented Architecture for Enterprise Applications"*, Wiley.
[6] Mark D. Hansen, *"SOA using Java Web Services"*, Practice Hall, 2007.
[7]WaseemRoshen, *"SOA-Based Enterprise Integration",* Tata McGraw-HILL, 2009.

BIOGRAPHIES AND PHOTOGRAPHS

1 *PUSHPALATHA B*

She is doing her Master of Technology (M Tech) degree in Computer Networking from New Horizon College of Engineering, Bangalore, Karnataka, India. She held the Bachelor of Technology (B Tech) degree in Computer Science and Engineering from BTL Institute of Technology and Management Bangalore, Karnataka, India. Her primary research interests are in the area of Big Data, Service Oriented Architecture, Cyber Security, and Information Network Security.

R.JAYA is working as a Sr.Assistant Professor in the department of CSE at New Horizon college of Engg.She has 13 years of teaching experience. She has done bachelor and master degree in engineering under Anna University. She is doing Ph.D under VTU. Her research interest includes data mining, expert systems,SOA and security.

An Efficient Reduction of Encryption Keys for Group Data Sharing via Cloud Storage

ManjulaK.[1],MahalakshmiG.[2],YashaswiniM.S[3]
Department of Information Science, Visveswaraya Technical University, Bangalore-60
manjukarur@gmail.com, mahalakshmi.gk25@gmail.com, yashu.cool968@mail.com

ABSTRACT-The art of selectively sharing encrypted data with different users via public clouds to rage may greatly ease security concerns with unintentional data leaks in the cloud. While sharing any group of selected documents with any group of users demands different encryption keys to be used for different documents.The user receives a large number of keys for both encryption and search. The user should also store the received keys and submit equal number of Keyword Secret door to cloud to perform search over the shared data. Due to the secure communication, complexity and problem, the above approach is impractical. So to address this problem,we propose a novel concept of One Key Search Many(OKSM) and instantiating the concept through a concrete One Key Search Many(OKSM) scheme,in which a data owner only needs to distribute a single key to a user for sharing a large number of documents, and the user only needs to submit a single Secretdoor to the cloud for querying the shared documents.

IndexTerms-Cloudstorage,Dataprivacy ,Datasharing, Searchable encryption.

I. INTRODUCTION

Nowadays the storage in the cloud has materialized as a capable answer for providing convenient and on-demand access to large amounts of data shared over the Internet. Today, millions of users are sharing personal data, such as photos and videos, with their friends through social network applications based on cloud storage on a daily basis. Business users are also being attracted by cloud storage due to its numerous benefits, including lower cost, greater agility, and better resource utilization.

However, while enjoying the convenience of sharing data via cloud storage, users are also concerned about inadvertent data leaks in the cloud. It is caused by a malicious adversary or a misbehaving cloud operator, can usually lead to serious breaches of personal privacy or business secrets.

To overcome users' concerns over potential data leaks in cloud storage, a common approach is for the data owner to encrypt all the data before uploading into the cloud such an cloud storage is often called the cryptographic cloud storage. However, the encryption of data makes it challenging for users to search and then selectively retrieve only the data containing given keywords. A common solution is to employ a searchable encryption (SE) scheme , in which the data owner is required to encrypt potential keywords and upload them to the cloud together with encrypted data, such that, for retrieving data matching a keyword, the user will send the corresponding keyword Secret door to the cloud for performing search over the encrypted data.

Although combining a searchable encryption scheme with cryptographic cloud storage can achieve the basic security requirements of a cloud storage, but not the efficient management of encryption keys. If the need for selectively sharing encrypted data with different users (e.g., sharing a photo with certain friends in a social network application, or sharing a business document with certain colleagues on a cloud drive) usually demands different encryption keys to be used for different files. However, this implies that the number of keys need to be distributed to users, for both to search over the encrypted files and to decrypt the files, will be proportional to the number of such files. Such a large number of keys must not only be distributed to users via secure channels, but also be securely stored and managed by the users in their devices.

Also a large number of secret doors must be generated by users and submitted to the cloud in order to perform a keyword search over many files. The secure communication, Complexity and storage tells that the above system is impractical.

We address this challenge by proposing the novel concept of One Key Search Many (OKSM), and instantiating the concept through a concrete OKSM scheme. The proposed OKSM scheme applies to any cloud storage that supports the searchable group data sharing functionality, which means any user may selectively share a group of selected files with a group of selected users. To support searchable group data sharing the main requirements for efficient key management are twofold.

As shown in the Fig.1,first, a data owner only needs to distribute a single One Key (instead of a group of keys) to a user for sharing any number of files. Second, the user only needs to submit a single secret door (instead of a group of secret doors) to the cloud for performing keyword search over any number of shared files.

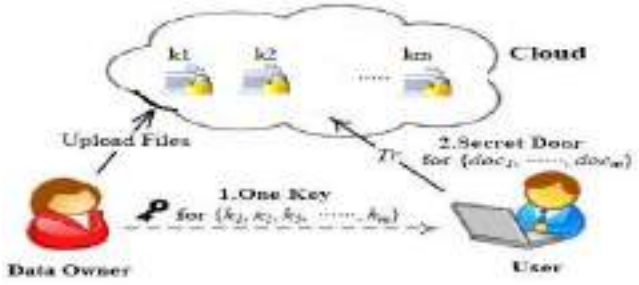


Fig.1: One Key Search Many

Our main contributions in this paper are as follows:

1.1 We first define a general framework of One Key Search Many (OKSM) composed of seven polynomial algorithms for security parameter setup, key generation, encryption, key extraction, secretdoor generation, secretdoor adjustment, and testing. We then describe both functional and security requirements for designing a valid OKSM scheme.

1.2 We then instantiate the OKSM framework by designing a concrete OKSM scheme. After providing detailed constructions for the seven algorithms, we analyze the efficiency of the scheme, and establish its security through detailed analysis.

1.3 We discuss various practical issues in building an actual group data sharing system based on the proposed OKSM scheme, and evaluate its performance. The evaluation confirms our system can meet the performance requirements of practical applications.

2. RELATED WORK

In Public Encryption With Keyword Search(PEKS) [1], the problem in public cloud system to search for encrypted data through encryption key is examined. Keyword as search query for email gateway is firstly introduced. Then based on the keyword the documents are routed to the gateways.PEKS system implies Identity Based Encryption (IDE) scheme where owner encrypts data such that user having required attributes can only decrypt the shared document. This system considered only single owner and user condition for performing keyword search over multiple shared documents.

In Symmetric Searchable Encryption(SSE)[2] ,one party allows to outsource the storage of its data to another party ,where another party is a server .In SSE there can be no sharing of data between two individuals, hence it is specified for a single user.

In Multi-user Searchable Encryption (MUSE) [3],[4]: It works under multi-tenancy operations where data owner shares documents with group of users and users can receive them by submitting trapdoor for keyword search on shared contents. That means users who have authorization can only retrieve the documents. It advances the single user SSE and PEKS schemes. But access control is not fine grained.

In Attribute Based Encryption(ABE) [5], narrowing the scope of search results to user's decryptable file's group can be done.ABE is of two types Key Policy(KP-ABE) andCipher Policy(CP-ABE) .CP- ABE isused to minimize the information leak and to reduce searching complexity when there are multi-users in cryptographic cloud storage. This system only search for related documents which user can decrypt and so is more efficient. The flexibility of specifying the access rights for individual users in case of user revocation is provided known as fine grained access control.

In Multi-Key Searchable Encryption(MKSE) [6],this system provides flexibility to user for searching over multiple documents which he/she can access say n, with different encryption keys. One search token is provided by user to server instead of n tokens. The user have to provide some public information and token for word to search and the system server then by using this information calculates token for different keys (adjust function) and get all documents with matching word even their encryption keys are different. Only single user and multi-key condition is considered in this system.

In Key Aggregate Encryption (KAE) for Data Sharing [7], the sharing of multiple documents with same user can be done , the data owner needed to distribute equal number of keys to the user. The complexity and security aspects become more subtle and impractical in case of large number of shared documents. In this scheme only single aggregate key to decrypt all documents is provided by owner i.e. compression of secret key. A public-key cryptosystem is introduced which produce constant-size cipher-texts leading to limited secure storage application. The user encrypts data under public key and ciphertext class which is identifier of ciphertext.

3 . IMPLEMENTATION

3.1 We first define the OKSM scheme ,which consist of seven algorithm .The algorithms are as follows :

3.1.1 **Setup**

This algorithm is run by the cloud service provider to set up the scheme. On input of a security parameter and the maximum possible number of documents which belongs to a data owner, it outputs the public system parameter Params.

3.1.2 **Keygen**

This algorithm is run by the data owner to generate a random key pair (pk,msk).

3.1.3 **Encipher**

This algorithm is run by the data owner to encrypt the i-th document and generate its keywords' ciphertexts. For each document, this algorithm will create a delta for its searchable encryption key ki. On input of the owner's public key pk and the file index i, this algorithm outputs data ciphertext and keyword ciphertexts Ci.

3.1.4 **Onegen**

This algorithm is run by the data owner to generate an OneKey ,searchable encryption key for delegating the keyword search right for a certain set of documents to other users. It takes as input the owner's master-secret key

msk and set S which contains the indices of documents ,then outputs OneKey.

3.1.5 **Secretdoor**

This algorithm is run by the user who has the onekey to perform a search. It takes as input the onekey searchable encryption key ok and a keyword w, then outputs only one secretdoor .

3.1.6 **Tune**

This algorithm is run by cloud server to adjust the combined secretdoor to generate the right secretdoor for each different document. It takes as input the system public parameters params, the set S of documents' indices, the index i of target document and the combined secretdoor Sr, then outputs each secretdoor Sri for the i-th target document in S.

3.1.7 **Search**

This algorithm is run by the cloud server to perform keyword search over an encrypted document. It takes as input the secretdoor and the document index i, then outputs true or false to denote whether the document contains the keyword w.

3.2 Designing of OKSM scheme :

For designing of OKSM scheme it should satisfy functional and security requirements , they are as follows:

Functional Requirements

- **Compactness:** The size of the One key has to be independent of the number of files to be shared .
- **Searchability:** This specifies the reduction of encryption has to preserve the keyword search.
- **Delegation :** The goal is to give the keyword search right to the user through One key.

Security Requirements

- **Controlled Searching:** Meaning that the intruder cannot search for an random word without the data owner's authorization.
- **Query Privacy:** Meaning that the intruder cannot determine the keyword used in the query because the user hides the actual keyword from the server also.

3.3 Evaluation of OKSM scheme

Here ,the various practical issues in building an actual group data sharing system based on the proposed OKSM scheme, and evaluate its performance . Through detailed analysis we achieve the functional and security requirements .

4. RESULTS

Our system is efficient as we can see that ,even when the number of documents increases which are to be shared ,the time cost is linear with the number of documents .It also shows the a single secretdoor will not reduce the efficiency over searching of document. This is showed in Fig.2,

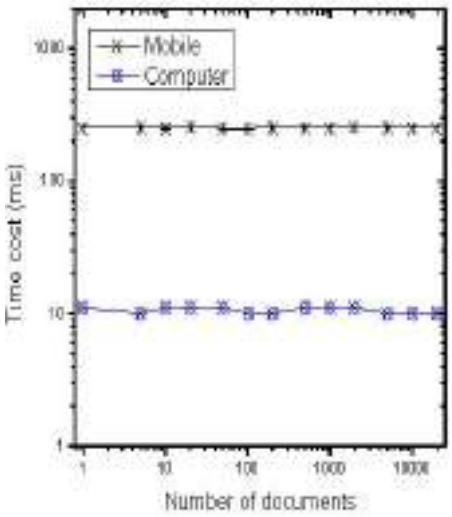


Fig.2: Time cost of secretdoor

5. CONCLUSION

Taking into consideration of the realistic problem of privacy preserving data sharing system based on public cloud storage which is need a data owner to allocate a large number of keys to users to permit them to access the documents, In this proposed concept of One Key Search Many (OKSM) and construct a concrete OKSM scheme. It can provide an efficient solution to building practical data sharing system based on public cloud storage. In a OKSM scheme, the owner needs to distribute a single key to a user when contributing a lot of documents with the user, and the user needs to submit a single secret door when they queries over all documents shared by the same owner.

On the other hand, if a user wants to question over documents shared by multiple owners, that user must produce multiple secret door to the cloud. The future enhancement for this proposed work is to find out how to decrease the number of secret door under multi-owners setting by attaining the security.

REFERENCES

Proceedings Papers:

[1] D. Boneh, C. G, R. Ostrovsky, G. Persiano. "Public KeyEncryption with Keyword Search" eurocrypt 2004, pp.506C522,2004.

[2] R. Curtmola, J. Garay, S. Kamara, R. Ostrovsky "Searchable Symmetric Encryption: improved definitions and efficient constructions", In: Proceedings of the 13th ACM conference on Computer

and Communications Security, ACM Press, pp. 79-88, 2006.

[3] Z. Liu, Z. Wang, X. Cheng, C. Jia and Ke Yuan, "Multi-user Searchable Encryption with Coarsed Grained Access Control in Hybrid Cloud", 2013.

[4] W. Li, J. Li, X. F. Chen, et al."Efficient` Keyword Search over Encrypted Data with Fine-Grained Access Control In Hybrid Cloud", In: Network and System Security 2012, LNCS, pp. 490-502, 2012.

[5] F. Zhao, T. Nishide, and K. Sakurai, "Multi- User Keyword Search Scheme for Secure Data Sharing with Fine-Grained Access Control", 2012.

[6] A. Popa and N. Zeldovich, "Multi-Key Searchable Encryption", 2013.

[7] C.Chu,S.Chow,W Tzeng,et al,"Key Aggregate Cryptosyste for Scalable data Sharing in Cloud Storage",IEEE Transaction parallel and distributed system ,2014.

Provision of Securing Data in Cloud through Auditing and Deduplication

Dhananjaya V, Naveen Ghorpade
Assistant Professor,CSE Department
SKIT,Bangalore
csdhananjay@gmail.com, naveen.ghorpade@gmail.com

Abstract— Outsourcing storage into the cloud is economically attractive for the cost and complexity of long-term large-scale data storage. At the same time, though, such a service is also eliminating data owners' ultimate control over the fate of their data, which data owners with high service-level requirements have traditionally anticipated. As owners no longer physically possess their cloud data, previous cryptographic primitives for the purpose of storage correctness protection cannot be adopted, due to their requirement of local data copy for the integrity verification. Besides, the large amount of cloud data and owner's constrained computing capabilities further makes the task of data correctness auditing in a cloud environment expensive and even formidable for individual cloud customers. Therefore, enabling secure audit-ability for cloud storage is of critical importance so that owners can resort to a secure cloud storage auditing and maintain strong storage correctness guarantee.

I. INTRODUCTION

Cloud storage provides customers with benefits, ranging from cost saving and simplified convenience, to mobility opportunities and scalable service. Even though cloud storage system has been widely adopted, it fails to accommodate some important emerging needs such as the abilities of auditing integrity of cloud files by cloud clients and detecting duplicated files by cloud servers. Even though cloud storage system has been widely adopted, it fails to accommodate some important emerging needs such as the abilities of auditing integrity of cloud files by cloud clients and detecting duplicated files by cloud servers. We illustrate both problems below.

The first problem is integrity auditing. The cloud server is able to relieve clients from the heavy burden of storage management and maintenance. These concerns originate from the fact that the cloud storage is susceptible to security threats from both outside and inside of the cloud [1], and the uncontrolled cloud servers may passively hide some data loss

Incidents from the clients to maintain their reputation. Considering the large size of the outsourced data files and the clients' constrained resource capabilities, the first problem is generalized as how can the client efficiently perform periodical integrity verifications even without the local copy of data files.

The second problem is secure deduplication. The rapid adoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers. This fact raises
a technology namely deduplication, in which the cloud servers would like to deduplicate by keeping only a single copy for each file (or block) and make a link to the file (or block) for every client who owns or asks to store the same file (or block).
Thus, the second problem is generalized as how can the cloud servers efficiently confirm that the client (with a certain degree assurance) owns the uploaded file (or block) before creating a link to this file (or block) for him/her.

In this paper, aiming at achieving data integrity and deduplication in cloud, we propose two secure systems namely
SecCloud and SecCloud+.

SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. SecCloud is supported on both block level and sector level. In addition, SecCloud also enables secure deduplication. A design a proof of ownership protocol between clients and cloud servers, which allows clients to prove to cloud servers that they exactly own the target data. Motivated by the fact that customers always want to encrypt their data before uploading, for reasons ranging from personal
privacy to corporate policy, we introduce a key server into SecCloud as with [4] and propose the SecCloud+ schema. SecCloud+ enables the guarantee of file confidentiality. The challenge of deduplication on encrypted is the prevention of dictionary attack [4]. As with [4], we make a modification on convergent encryption such that the convergent key of file is generated and controlled by a secret "seed", such that any adversary could not directly derive the convergent key from the content of file and the dictionary attack is prevented.

II. LITERATURE SURVEY

We review the works in both areas in the following subsections, respectively.

A. Integrity Auditing

The definition of provable data possession (PDP) was introduced by Ateniese et al. [5] [6] for assuring that the cloud servers possess the target files without retrieving or downloading the whole data. PDP is a probabilistic proof protocol by sampling a random set of blocks and asking the servers to prove that they exactly possess these blocks, and

the verifier only maintaining a small amount of metadata is able to perform the integrity checking. Ateniese et al. [7] proposed a dynamic PDP schema but without insertion operation; Erway et al. [8] improved Ateniese et al.'s work [7] and supported insertion by introducing authenticated flip table; A similar work has also been contributed in [9]. Wang et al. [10] proposed proxy PDP in public clouds. Zhu et al. [11] proposed
the cooperative PDP in multi-cloud storage. Line of work supporting integrity auditing is proof of retrievability (POR) [12].
Compared with PDP, POR not merely assures the cloud servers possess the target files, but also guarantees their full recovery. In [12], clients apply erasure codes and generate authenticators for each block for verifiability and retrievability. Wang et al. [13] improved the POR model by manipulating the classic Merkle hash tree construction for block tag authentication. Xu and Chang [14] proposed to improve the POR schema in [12] with polynomial commitment for reducing communication cost. Stefanov et al. [15] proposed a POR protocol over authenticated file system subject to frequent changes. Azraoui et al. [16] combined the privacy-preserving word search algorithm with the insertion in data segments of randomly generated short bit sequences, and developed a new POR protocol. Li et al. [17] considered a new cloud storage architecture with two independent cloud servers for integrity auditing to reduce the computation load at client side. Recently, Li et al. [18] utilized the key-disperse paradigm to fix the issue of a significant number of convergent keys in convergent encryption.

B. Secure Deduplication

Deduplication is a technique where the server stores only a single copy of each file, such that the disk space of cloud servers as well as network bandwidth are saved. Trivial client side deduplication leads to the leakage of side channel information. In order to restrict the leakage of side channel information, Halevi et al. [3] introduced the proof of ownership protocol which lets a client efficiently prove to a server that that the client exactly holds this file. Pietro and Sorniotti [19] proposed an efficient proof of ownership scheme by choosing the projection of a file onto some randomly selected bit-positions as the file proof.
Line of work for secure deduplication focuses on the confidentiality of deduplicated data and considers to make deduplication on encrypted data. Ng et al. [20] firstly introduced the private data deduplication as a complement of public data deduplication protocols of Halevi et al. [3]. Convergent encryption [21] is a promising cryptographic primitive for ensuring data privacy in deduplication. Bellare et al. [22] formalized this primitive as message-locked encryption, and explored its application in space-efficient secure outsourced storage. Abadi et al. [23] further strengthened Bellare et al's security definitions [22] by considering plaintext distributions that may depend on the public parameters of the schemas. Regarding the practical implementation of convergent encryption for securing deduplication, Keelveedhi et al. [4] designed the DupLESS system in which clients encrypt under file-based keys derived from a key server via an oblivious pseudorandom function protocol.

III. ALGORITHMS USED AND COMPUTATION

Some preliminary notions that will form the foundations of our approach.

A. Convergent Algorithm

Convergent encryption [22][23][21] provides data confidentiality in deduplication. A user (or data owner) derives a convergent key from the data content and encrypts the data copy with the convergent key. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property [22] holds, i.e., if two data copies are the same, then their tags are the same. Formally, a convergent encryption scheme can be defined with four primitive functions:

- **KeyGen(F)** : The key generation algorithm takes a file content F as input and outputs the convergent key ckF of F.
- **Encrypt($ckF;F$)** : The encryption algorithm takes the convergent key ckF and file content F as input and outputs the ciphertext ctF.
- **Decrypt($ckF; ctF$)** : The decryption algorithm takes the convergent key ckF and ciphertext ctF as input and outputs the plain file F.
- **TagGen(F)** : The tag generation algorithm takes a file content F as input and outputs the tag $tagF$ of F. Notice that in this paper, we also allow TagGen(·) to generate the (same) tag from the corresponding ciphertext as with [22][18].

B. Bilinear Map and Computational Assumption

Definition 1 (Bilinear Map): Let G and GT be two cyclic multiplicative groups of large prime order p. A bilinear pairing is a map $e : G \times G \rightarrow GT$ with the following properties:

- **Bilinear:** $e(ga1 ; gb2) = e(g1; g2)ab$ for all $g1; g2 \in_R G$ and $a; b \in_R Zp$.
- **Non-degenerate:** There exists $g1; g2 \in G$ such that $e(g1; g2) \neq 1$.
- **Computable:** There exists efficient algorithm to compute $e(g1; g2)$ for all $g1; g2 \in_R G$.

The examples of such groups can be found in supersingular elliptic curves or hyperelliptic curves over finite fields, and the
bilinear pairings can be derived from the Weil or Tate pairings. For more details, see [24]. We then describe the Computational Diffie-Hellman problem, the hardness of which will be the basis of the security of our proposed schemes.

Definition 2 (CDH Problem): The Computational Diffie-Hellman problem is that, given $g; gx; gy \in G1$ for unknown $x; y \in Z*p$, to compute gxy.

IV. SECCLOUD

In this section, we describe our proposed SecCloud system. we begin with giving the architecture of SecCloud as well as introducing the design goals for SecCloud.
Aiming at allowing for auditable and deduplicated storage,we propose the SecCloud system. In the SecCloud system, we have three entities:

• Auditor which helps clients upload and audit their outsourced data maintains a Map Reduce cloud and acts like a certificate authority.
• Cloud Servers virtualize the resources according to the requirements of clients and expose them as storage pools.
• Cloud Clients have large data files to be stored and rely on the cloud for data maintenance and computation. They can be either individual consumers or commercial Organizations.

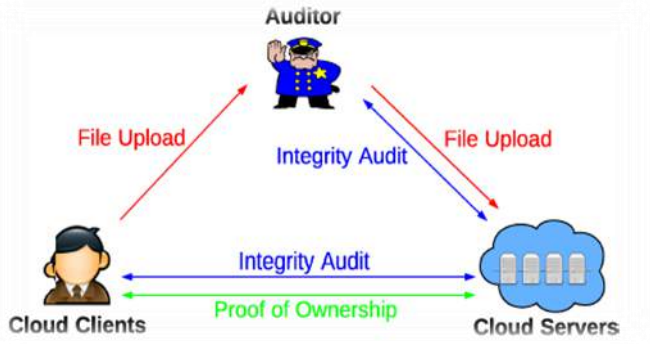


Fig. 1. SecCloud Architecture

The SecCloud system supporting file-level deduplication includes the following three protocols respectively highlighted in Fig. 1.

File Uploading Protocol: This protocol aims at allowing clients to upload files via the auditor. Specifically, the file uploading protocol includes three steps:
• Step 1: (cloud client → cloud server): client performs the duplicate check with the cloud server to confirm if such a file is stored in cloud storage or not before uploading a file. If there is a duplicate, another protocol called Proof of Ownership will be run between the client and the cloud storage server. Otherwise, the following protocols (including step 2 and step 3) are run between these two entities.
• Step 2 (cloud client → auditor): client uploads files to the auditor, and receives a receipt from auditor.
• Step 3 (auditor → cloud server): auditor helps generate a set of tags for the uploading file, and send them along with this file to cloud server.

Integrity Auditing Protocol: It is an interactive protocol for integrity verification and allowed to be initialized by any entity except the cloud server. In this protocol, the cloud server plays the role of prover, while the auditor or client works as
the verifier. This protocol includes two phases:
• Step 1 (cloud client/auditor → cloud server): verifier (i.e., client or auditor) generates a set of challenges and sends them to the prover (i.e., cloud server).
• Step 2 (cloud server → cloud client/auditor): based on the stored files and file tags, prover (i.e., cloud server) tries to prove that it exactly owns the target file by sending the proof back to verifier (i.e., cloud client or auditor).
At the end of this protocol, verifier outputs true if the integrity verification is passed.

Proof of Ownership Protocol: It is an interactive protocol initialized at the cloud server for verifying that the client exactly owns a claimed file. This protocol is typically triggered along with file uploading protocol to prevent the leakage of side channel information. On the contrast to integrity auditing protocol, in PoW the cloud server works as verifier, while the client plays the role of prover. This protocol also includes two steps:
• Step 1 (cloud server → client): cloud server generates a set of challenges and sends them to the client.
• Step 2 (client → cloud server): the client responds with the proof for file ownership, and cloud server finally verifies the validity of proof.
Our main objectives are outlined as follows:
• **Cost-Effective**: The computational overhead for providing integrity auditing and secure deduplication should not represent a major additional cost to traditional cloud storage, nor should they alter the way either uploading or downloading operation.
• **Integrity Auditing**: The first design goal of this work is to provide the capability of verifying correctness of the remotely stored data. The integrity verification further requires two features: 1) *public verification*, which allows anyone, not just the clients originally stored the file, to perform verification; 2) *stateless verification*, which is able to eliminate the need for state information maintenance at the verifier side between the actions of auditing and data storage.
• **Secure Deduplication**: The second design goal of this work is secure deduplication. In other words, it requires that the cloud server is able to reduce the storage space by keeping only one copy of the same file. Notice that, regarding to secure deduplication, our objective is distinguished from previous work [3] in that we propose a method for allowing both deduplication over files and tags.

B. SecCloud Details

In this subsection, we respectively describe the three protocols including file uploading protocol, integrity auditing protocol and proof of ownership protocol in SecCloud. Before our detailed elaboration, we firstly introduce the system setup phase of SecCloud, which initializes the public and private parameters of the system.
• **System Setup:** The auditor working as an authority picks a random integer $\_ \in_R Zp$ as well as random elements $g; u1; u2; : : : ut \in_R G$, where t specifies the maximum number of sectors in a file block. The secret key sk is set to be $\_$ and kept secret, while the public key $pk = (g\_; \{ui\}t\ i{=}1)$ is published to other entities.
File Uploading Protocol: Based on the public and private parameters generated in system setup, we then describe the file uploading protocol. Suppose the uploading file F has s blocks: $B1;B2; : : : ;Bs$, and each block Bi for $i = 1; 2; : : : ;$ s contains t sectors: $Bi1;Bi2; : : : ;Bit$. Let n be the number of slave nodes in the MapReduce cloud. The client runs the deduplication test by sending hash value of the file Hash(F) to the cloud server. If there is a duplicate, the cloud client performs Proof of Ownership protocol with the cloud server which will be described later. If it is passed, the user is authorized to access this stored file without uploading the file.
Otherwise (in the second phase), the cloud client uploads a file F as well as its identity IDF to the distributed file system in MapReduce auditing cloud, and simultaneously sends an "upload" request to the master node in MapReduce, which randomly picks $\{\_i\}ni\ {=}1$ such that $\Sigma n\ i{=}1\ \_i = \_$ and assigns the ith slave node with $\_i$. When each slave node (say the ith

salve node) receives the assignment _*i*, it does two steps: 1) Pick up (ID*F;F*) in the distributed file system in MapReduce, and build a Merkle hash tree on the blocks *{Bj}sj*=1 of *F*. 2)
Let *hroot* denote the hash of the root node of Merkle hash tree
built on *F*. This slave node uses _*i* to sign *hroot* by computing
_*i* = *h_iroot*. Finally, the signature _*i* is sent to the the slave node which is specified by master node for executing the *reducing* procedure. The specified slave node for reducing procedure gathers all the signatures *{_i}ni* =1 from the other slave nodes, and computes _ = Π*n i*=1 _*i*. The "reduced" signature _ is finally sent back to client as receipt of the storage of file *F*. In the third phase, the MapReduce auditing cloud starts to upload the file *F* to cloud server. To allow public auditing, the master node builds file tags of *F*. Specifically, master node firstly writes and arranges all the sectors of *F* in a matrix (we say *S*), and computes a homographic signature for each *row* of the matrix *S* (highlighted red in Fig. 3). Notice that the tag generation procedure also follows the computing paradigm with MapReduce. That is, for the *i*th (*i* = 1*; 2; : : : ; s*) row of *S*, the *j*th (*j* = 1*; 2; : : : ; n*) slave node computes _*ij* = [Hash(ID*F*||*Bi*) Π*t k*=1 *uBik k*]_*j* , where Σ*n j*=1 _*j* = _. Accordingly, all the signatures *{_ij}nj* =1 are then multiplied into the homomorphic signature _*i* = Π*n j*=1 _*ij* at a specified reducing slave node. The homomorphic signature allows us to in future aggregate the signatures signed on the sectors in the same *column* of *S* using multiplication. Finally, the master node uploads (ID*;F; {_i}si* =1) to cloud server.

Integrity Auditing Protocol: In the integrity auditing protocol, either the MapReduce auditing cloud or the client works as the verifier. Thus, without loss of generality, in the rest of the description of this protocol, we use verifier to identify the client or MapReduce auditing cloud. The auditing protocol is designed in a challenge-response model. Specifically, the verifier randomly picks a set of block identifiers (say *IF*) of *F* and asks the cloud server (working as prover) to response the blocks corresponding to the identifiers in *IF*. In order to keep randomness in each time of challenge, even for the same *IF*, we introduce a random coefficient for each block in challenge. That is, for each identifier *i* ∈ *IF*, the coefficient *ci* for the block identified by *i* is computed as *ci* = *f*(*tm*||ID*F*||*i*), where *f*(·) is a pseudorandom function and *tm* is the current time period. Finally, *C* = *{*(*i; ci*)*}i*∈*IF* is sent to cloud server for challenge.

Proof of Ownership Protocol: The PoW protocol aims at allowing secure deduplication at cloud server. Specifically, in deduplication, a client claims that he/she has a file *F* and wants to store it at the cloud server, where *F* is an existing file having been stored on the server. The cloud server asks for the
proof of the ownership of *F* to prevent client unauthorized or malicious access to an unowned file through making cheating
claim. In SecCloud, the PoW protocol is similar to [3] and the
details are described as follows. Suppose the cloud server wants to ask for the ownership proof for file *F*. It randomly picks a set of block identifiers, say *IF* ⊆ *{*1*; 2; : : : ; s}* where *s* is the number of blocks in *F*, for challenge. Upon receiving the challenge set *IF*, the client first computes a short value and constructs a Merkle tree. Note that only sibling-paths of all the leaves with challenged identifiers are returned back to the cloud server, who can easily verify the correctness by only using the root of the Merkle tree. If it is passed, the user is authorized to access this stored file.

V. SECCLOUD+

We specify that our proposed SecCloud system has achievedboth integrity auditing and file deduplication. However, itcannot prevent the cloud servers from knowing the content of files having been stored. In other words, the functionalitiesof integrity auditing and secure deduplication are only imposed on plain files. In this section, we propose SecCloud+, which allows for integrity auditing and deduplication on encrypted files.

A. System Architecture

Compared with SecCloud, our proposed SecCloud+ involves an additional trusted entity, namely key server, which is responsible for assigning clients with secret key (according to the file content) for encrypting files. This architecture is in line with the recent work [4]. But our work is distinguished with the previous work [4] by allowing for integrity auditing on encrypted data. SecCloud+ follows the same three protocols (i.e., the file uploading protocol, the integrity auditing protocol and the proof of ownership protocol) as with SecCloud. The only difference is the file uploading protocol in SecCloud+ involves an additional phase for communication between cloud client and key server. That is, the client needs to communicate with the key server to get the convergent key for encrypting the uploading file before the phase 2 in SecCloud.
Unlike SecCloud, another design goals of file confidentiality is desired in SecCloud+ as follows.

• **File Confidentiality**: The design goal of file confidentiality requires to prevent the cloud servers from accessing the content of files. Specially, we require that the goal of file confidentiality needs to be resistant to "dictionary attack". That is, even the adversaries have pre-knowledge of the "dictionary" which includes all the possible files, they still cannot recover the target file [4].

B. SecCloud+ Details

We introduce the system setup phase of SecCloud+ as follows.

• **System Setup:** As with SecCloud, the auditor initializes the public key *pk* = (*g_; {ui}t i*=1) and private key *sk* = _, where *g; u*1*; u*2*; : : : ; ut* ∈*R* G. In addition, to preserve the confidentiality of files, initially, the key server picks a random key *ks* for further generating file encryption keys, and each client is assigned with a secret key *ck* for encapsulating file encryption keys. Based on the initialized parameters, we then respectively describe the three protocols involved in SecCloud+.

File Uploading Protocol: Suppose the uploading file *F* has *s* blocks, say *B*1;*B*2*; : : : ;Bs*, and each block *Bi* for *i* = 1*;* 2*; : : : ; s* contains *t* sectors, say *Bi*1;*Bi*2*; : : : ;Bit*.Client computes *hF* = Hash(*F*) by itself. In addition, for each sector *Bij* of *F* where *i* = 1*;* 2*; : : : ; s* and *j* = 1*;* 2*; : : : ; t*, client computes its hash *hBij* = Hash(*Bij*). Finally (*hF; {hBi}i*=1;*:::;s;j*=1;*:::;t*) is sent to key server for generating the convergent keys for *F*.

Upon receiving the hashes, the key server computes *sskF* = *f*(*ks; hF*) and *sskij* = *f*(*ks; hBij*) for *i* = 1*; : : : ; s* and *j* = 1*; : : : ; t*, where *ks* is the convergent key seed kept at the key server, and *f*(·) is a pseudorandom function. It is worthwhile nothing that, 1) We take advantage of the idea of convergent encryption [21][22][23] to make the deterministic and "content identified" encryption, in which each "content" (file or sector) is encrypted using the session key derived from itself. In this way, different "contents" would result in different ciphertexts, and deduplication works. 2) Convergent encryption suffers from dictionary attack, which allows the adversary to recover the whole content with a number of guesses. To prevent such attack, as with [4], a "seed" (i.e., convergent key seed) is used for controlling and generating all the convergent keys to avoid the fact that adversary could guess or derive the convergent key just from the content itself. 3) We generate convergent keys on sector-level (i.e., generate convergent keys for each sector in file *F*), to enable integrity auditing. Specifically, since convergent encryption is deterministic, it allows to compute homomorphic signatures on (convergent) encrypted data as with on plain data, and thus the sector-level integrity auditing is preserved. Client then continues to encrypt *F* sector by sector and uploads the ciphertext to auditor. Specifically, for each sector *Bij* of *F*, *i* = 1*;* 2*; : : : ; s* and *j* = 1*;* 2*; : : : ; t*, client computes *ctBij* = Enc(*sskBij ;Bij*), and sends (ID*F; {ctBij }i*=1;*:::;s;j*=1;*:::;t*) to auditor, where Enc(·) is the symmetric encryption algorithm. The convergent keys *sskij* are encapsulated by client's secret key *ck* and directly stored at the cloud servers.

Integrity Auditing Protocol: The integrity auditing protocol works in the same way of that in SecCloud, but imposed on encrypted data. Specifically, the verifier (could be either the client or the auditor) submits a set of pairs *{*(*i; ci*)*}i∈IF* where *IF* ⊆ *{*1*;* 2*; : : : ; s}* and *ci* ∈*R* Z. Upon receiving *{*(*i; ci*)*}i∈IF* , the cloud servers then computes *!j* = Σ *i∈IF cictBij* for each

j = 1*;* 2*; : : : ; t*, as well as the aggregated homomorphic signature _ = Π *i∈IF* _*ci i* . In addition, the cloud server constructs a Merkle hash tree on encrypted blocks *ctBi* of *F* and attempts to prove retrievability at block-level. Precisely, for each *i* ∈ *IF*, the cloud server computes a pair (Hash(*ctBi*);Ω*i*), where *ctBi* = [*ctBi*1 *; : : : ; ctBit*] and Ω*i* includes the necessary auxiliary information for reconstructing the root node using *{ctBi }i∈IF* . Finally (_*; {!j}t j*=1*; {*(Hash(*ctBi*);Ω*i*)*}i∈IF*) is sent to verifier for auditing.

VI. SECURITY ANALYSIS

In this section, we attempt to analyze the security of our proposed both schemes. Before this, we firstly formalize the security definitions our schemes aim at capturing.

A. Security Definitions

Based on the paradigm of SecCloud and SecCloud+, we define the security definitions, adapting to the integrity auditing and secure deduplication goals. Our both definitions capture the philosophy of game-based definition. Specifically, we define two games respectively for integrity auditing and secure deduplication, and both of the games are played by two players, namely adversary and challenger. The adversary (the role of which is worked by semi-honest cloud server and cloud client respectively in integrity auditing and secure deduplication definition) is trying to achieve the goal condition explicitly specified in the game. Having this intuition, we give our security definitions as follows.

1) Integrity Auditing: An integrity auditing protocol is sound if any cheating cloud server that convinces the verifier that it is storing a file *F* is actually storing this file. To capture this spirit, we define its game based on Proof of Retrievability (PoR). The security model called Proof of Retrievability (PoR) was introduced by Shacham and Waters' in [12]. This security model captures the requirement for integrity auditing, whose basic security goal is to achieve proof of retrievability. In more details, in this security model, if there exists an adversary who can forge and generate any valid integrity proofs for any file *F* with a non-negligible probability, another simulator can be constructed who is able to extract *F* with overwhelming probability. The formal definition for the above model can be given by the following game between a challenger and an adversary *A*. Note that in the following security game, the challenger plays the role of auditing server while the adversary *A* acts as the storage server.

• **Setup Phase**: The challenger runs the setup algorithm with required security parameter and other public parameter as input. Then, it generates the public and secret key pair (*pk; sk*). The public key *pk* is forwarded to the adversary *A*.

• **Query phase**: The adversary is allowed to query the file upload oracle for any file *F*. Then, the file with the correct tags are generated and uploaded to the cloud storage server. These tags can be publicly verified with respect to the public key *pk*.

• **Challenge Phase**: *A* can adaptively send file *F* to the file tag *tag* comes, *C* runs the integrity verification protocol IntegrityVerify*{A C*(*pk; tag*)*}* with *A*.

• **Forgery**: *A* outputs a file tag *tag′* and the description of a prover *Pt*.

Secure Deduplication: Similarly, we can also define a game between challenger and adversary for secure deduplication below. Notice that the game for secure deduplication captures the intuition of allowing the malicious client to claim it has a challenge file *F* through colluding with all the other clients not owning this file.

• **Setup Phase**: A challenge file *F* with fixed length and minimum entropy (specified in system parameter) is randomly picked and given to the challenger. The challenger continues to run a summary algorithm and generate a summary *sumF*.

• **Learning Phase**: Adversary *F* can setup arbitrarily many client accomplices not exactly having *F* and have them to

interact with the cloud servers to try to prove the ownership of file *F*. Notice that in the learning phase, the cloud server plays as the honest verifier with input sum *sumF* and the accomplices could follow any arbitrary protocol set by *A*.

• **Challenge Phase**: The exact proof of ownership protocol is executed. Specifically, the challenger outputs a challenge to *A* and *A* responses with a proof based on its learnt knowledge. If *A*'s proof is accepted by the cloud server, we say *A* succeeds.

VII. PERFORMANCE ANALYSIS

In this section, we will provide a thorough experimental evaluation of our proposed schemes. We build our test bed by
using 64-bit t2.Micro Linux servers in Amazon EC2 platform as the auditing server and storage server. In order to achieve _ = 80 bit security, the prime order *p* of the bilinear group G and G*T* are respectively chosen as 160 and 512 bits in length. We also set the block size as 4 KB and each block includes 25 sectors. It is clear the time cost of slave node is growing with the size of file. This is because the more blocks in file, the more homomorphic signatures are needed to be computed by slave node for file uploading. We also need to notice that there does not exist much computational load difference between common slave nodes and the reducer. Compared with the common slave nodes, reducer only additionally involves in a number of multiplications, which is lightweight operation. It is worthwhile noting that, the procedure of tag generation (the phase 2 and 3 in file uploading protocol) could be handled in preprocessing, and it is not necessary for client to wait until uploading file.

Before examine the time cost of file auditing, we need to firstly make analysis and identify the number of challenging blocks (i.e., |*IF*|) in our integrity auditing protocol. According to [5], if _ fraction of the file is corrupted, through asking the proof of a constant *m* blocks of this file, the verifier can detect the misbehavior with probability _ = 1 − (1 − _)*m*. To capture the spirit of probabilistic auditing, we set the probability confidence _ = 70%*;* 85% and 99%, and draw the relationships between _ and *m* in Fig. 6. It demonstrates that if we want to achieve low (i.e., 70%), medium (i.e., 85%) and high (i.e., 99%) confidence of detecting any small fraction of corruption, we have to respectively ask for 130*;* 190 and 460 blocks for challenge.

VIII. CONCLUSION

Aiming at achieving both data integrity and deduplication in cloud, we propose SecCloud and SecCloud+. SecCloud introduces an auditing entity with maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. In addition, SecCoud enables secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading, and allows for integrity auditing and secure deduplication directly on encrypted data.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (No.61100224 and No. 61472091), NSFCGuangdong (U1135002) and Natural Science Foundation of Guangdong Province (Grant No. S2013010013671).

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communication of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in *IEEE Conference on Communications and Network Security (CNS)*, 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*. ACM, 2011,
pp. 491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in *Proceedings of the 22Nd USENIX Conference on Security*, ser. SEC'13. Washington,
D.C.: USENIX Association, 2013, pp. 179–194. [Online]. Available:https://www.usenix.org/conference/usenixsecurity 13/technicalsessions/presentation/bellare

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598– 609.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 12:1–12:34, 2011.

[7] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in *Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks*, ser. SecureComm '08. New York, NY, USA: ACM, 2008, pp. 9:1–9:10.

[8] C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 213–222.

[9] F. Seb´e, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," *IEEE Trans. on Knowl. and Data Eng.*, vol. 20, no. 8, pp. 1034–1038, 2008.

[10] H. Wang, "Proxy provable data possession in public clouds," *IEEE Transactions on Services Computing*, vol. 6,

no. 4, pp. 551–559, 2013. [11] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data
possession for integrity verification in multicloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2231– 2244, 2012.
[12] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology*, ser. ASIACRYPT '08. Springer Berlin Heidelberg, 2008, pp. 90–107.
[13] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in *Computer Security – ESORICS 2009*, M. Backes and P. Ning, Eds., vol. 5789. Springer Berlin Heidelberg, 2009, pp. 355–370.
[14] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," in *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '12. New York, NY, USA: ACM, 2012, pp. 79–80.
[15] E. Stefanov, M. van Dijk, A. Juels, and A. Oprea, "Iris: A scalable cloud file system with efficient integrity checks," in *Proceedings of the 28th Annual Computer Security Applications Conference*, ser. ACSAC '12. New York, NY, USA: ACM, 2012, pp. 229–238.
[16] M. Azraoui, K. Elkhiyaoui, R. Molva, and M. O¨ nen, "Stealthguard: Proofs of retrievability with hidden watchdogs," in *Computer Security - ESORICS 2014*, ser. Lecture Notes in Computer Science, M. Kutyłowski and J. Vaidya, Eds., vol. 8712. Springer International Publishing, 2014, pp. 239–256.
[17] J. Li, X. Tan, X. Chen, and D. Wong, "An efficient proof of retrievability with public auditing in cloud computing," in *5th International Conferenc on Intelligent Networking and Collaborative Systems (INCoS)*, 2013, pp. 93–98.
[18] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1615–1625, June 2014.
[19] R. Di Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for deduplication," in *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '12. New York, NY, USA: ACM, 2012, pp. 81–82.
[20] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, ser. SAC '12. New York, NY, USA: ACM, 2012, pp. 441–446.
[21] J. Douceur, A. Adya, W. Bolosky, P. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in *22nd International Conference on Distributed Computing Systems*, 2002, pp. 617–624.
[22] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in *Advances in Cryptology – EUROCRYPT 2013*, ser. Lecture Notes in Computer Science, T. Johansson and P. Nguyen, Eds. Springer Berlin Heidelberg, 2013, vol. 7881, pp. 296–312.
[23] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lock-dependent messages," in *Advances in Cryptology – CRYPTO 2013*, ser. Lecture Notes in Computer Science, R. Canetti and J. Garay, Eds. Springer Berlin Heidelberg, 2013, vol. 8042, pp. 374–391.
[24] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Advances in Cryptology — CRYPTO 2001*, ser. Lecture Notes in Computer Science, J. Kilian, Ed. Springer Berlin Heidelberg, 2001, vol. 2139, pp. 213–229.

A THIRD GENERATION DESIGN FOR SECURED ATM TRANSACTION USING SIM CARD

A.Chandhanaa , Durga.S
Adhiyamaan college of engineering, Hosur, Dept., Of CSE
chandhanaa95@gmail.com,durga.hosur@gmail.com

ABSTRACT- The use of mobile handheld devices is expanding rapidly both within the business and individual context. These devices are now essential tools that offer competitive business advantages in today's growing world of ubiquitous computing environment. The technology advancement has made it possible to embed more facilities in mobile phones. While they provide benefits, they also pose new risks on security either by the information they contain or information that they can access remotely. Secure cash transaction is of serious concern in growing use of cash cards and internet transactions. The main objective of this project is to create a new generation of ATM machine in which transaction can be made without ATM cards. Because in existing system the user should carry the ATM card without fail. In this system ATM machine can be operated by using SIM in the mobile phone. When the SIM is inserted in the reader unit of the ATM machine it transfers the mobile number information to the server. In server the related information of the mobile number (i.e) the users account details, their photos etc. are collected. The Face recognition module and Fingerprint recognition module present in the ATM machine compares the image and fingerprint ridges of the consumers with the existing one in the server. Only when the image and fingerprint matches, it asks for the PIN number and further processing starts. Otherwise the process is terminated. So, by using this system the need of ATM card is completely eliminated. We can operate the ATM machine with SIM itself. This paper will introduce the concept of physical browsing and development of a system that will allow users to use their mobile phones to securely withdraw cash from ATM machines. The paper presents the new generation of ATM machine for M-Cash withdrawal application, relevant technologies and security issues. By using this system malfunctions can be avoided and the transaction will be much secured.

1. INTRODUCTION

Automated teller machine is a mechanical device that has its roots embedded in the accounts and records of a banking institution. It is a machine that allows the bank customers to carry out banking transactions like, deposits, transfers, balance enquiries, mini statement, withdrawal and fast cash etc. Notwithstanding, we lived in a world where people no longer want to encounter long queues for any reason, they don't not want to wait for too long time before they are attended to and this has led to the increasing services being rendered by banks to further improve the convenience of banking through the means of electronic banking. On this note the advent of ATM is imperative, although with its own flaws. Crime at ATM's has become a nationwide issue that faces not only customers, but also bank operators. Security measures at banks can play a critical, contributory role in preventing attacks on customers. These measures are of paramount importance when considering vulnerabilities and causation in civil litigation and banks must meet certain standards in order to ensure a safe and secure banking environment for their customers. Basically, the ATM scam involves thieves putting a thin, clear, rigid plastic sleeve into the ATM card slot. When you insert your card, the machine can't read the strip, so it keeps asking you to re-enter your PIN number. Meanwhile, someone behind you watches as you tap in your number. Eventually you give up, thinking the machine has swallowed your card and you walk away. The thieves then remove the plastic sleeve complete with card, and empty your account. The main fact that many of the customers have never used an ATM before and are completely unfamiliar with that concept therefore they are very unlikely to memorize and remember a PIN. Furthermore, there is a sense of mistrust with PINs. People may feel that it is unsafe because if they lose their card they worry that someone will find and somehow be able to determine their PIN and steal their money from the ATM. To keep it in mind we proposed a combined technique i.e. costumers insert their card & PIN, if costumers insert valid PIN then access is grant to another security approved process i.e. biometric fingerprint and face recognition. Using valid PIN, fingerprint and image of the costumer can access ATM transaction process i.e. deposits, transfers, balance enquiries, mini statement, Fast cash & withdrawal etc. By using fingerprint recognition customers are more comfortable with the idea of saving their money with the bank because they understand that if they lose their ATM card, no one can replicate their fingerprint and take their money. The way to avoid this is to run your finger along the card slot before you put your card in. The sleeve has a couple of tiny prongs that the thieves need to get the sleeve out of the slot, and you'll be able to feel them. The primary focus of this work is on developing a biometric strategy (Fingerprint) to enhance the security features of the ATM for effective banking transaction and more comfortable feature i.e. we proposed another option for nominee user because in case a card holder faces an accident, then the transactions process is not possible. To keep this drawback in mind we consider nominees fingerprint sample for second user to do the transaction while actual card holder unable to do the transactions. Actually PIN code are changeable but fingerprint and image of the person are not changeable, so card holder may changes his/her PIN code while maintaining one's own secrecy and may permit his/her nominee with giving updated PIN code for transactions. We have considered the left & right thumb impression of an individual; it has been observed that there is no any match in these samples in any case. We have also observed that

thumb impression samples have been taken in different angles & different forces. To achieve various functionality and make good use of mobile devices so that people can keep them and utilized their potentials.

2. EXISTING BANKING ATM SYSTEM IN INDIA

There is no doubt that rapid development of banking technology has changed the way in dealing with banking activities. One of the examples is automatic teller machine (ATM). Using ATM, a customer is able to conduct several banking activities such as cash withdrawal, money transfer, paying phone and electricity bills beyond official hours and physical interaction with bank staff. In short, ATM provides customers a quick and convenient way to access their bank accounts and to conduct financial transactions. Password or personal identification number (PIN) is one of important aspects in ATM security system which is commonly used to secure and protect financial information of customers from unauthorized access. The system compares the code against a stored list of authorized passwords and users. PIN typically in a form of four digit combination of numbers that entered through ATM panel. If the code is legitimate, the system allows access at the security level approved for the owner of the account. In general, PIN is sufficient to protect against fraud and effectively eliminating most common attempts to gain unauthorized access. The four digit PIN is also easy to memorize and can be typed quickly with few errors and is quite difficult to be cracked if it is managed properly. The most recent cases show that the thefts have used sophisticated cracking programs to steal ATM holders money very easily, some people who live in today's high tech society which are bombarded everyday by so many numbers such as social security number, computer password, credit card number and so on. Sometimes they are confusing, difficult to be recalled immediately which of course can lead to a serious problem. Sometimes it is written down on small piece of paper or on ATM card in order to anticipate such event. The strength of PIN as a security system is weakened since the likelihood of the code leaking to other people increased. A personal identification number (PIN) can be +9used in much the same as a password. It is numerical in format and like a password that should be kept secret. The most common use of the PIN is in automatic teller machines (ATM). "Most commonly PINs are 4-digit numbers in the range 0000-9999 resulting in 10,000 possible numbers, so that an attacker would need to guess an average of 5000 times to get the correct PIN." Biometrics is a rapidly evolving technology that is being widely used in forensics, such as criminal identification and prison security, and that has the potential to be used in a large range of civilian application areas. Biometrics can be used to prevent unauthorized access to ATMs, cellular phones, smart cards, desktop PCs, workstations, and computer networks. There has been a growing use of Radio Frequency Identification Tags (RFID) in different business environments. A typical example includes supermarkets, airline industry and the majority of supply chains. The main advantage of RFID to businesses is being the effectiveness of identifying System that transmits the identity (in the form of a unique serial number) of an object wirelessly, using radio waves luggage/product; providing good customer service; cost reduction and flexibility. In this paper this technology has been used in a new application called Secure M-Cash Withdrawal, a system whereby a mobile phone is equipped with RFID tags and NFC. The mobile will interact with an ATM machine which is equipped with RFID reader and writer and a banking system. The process of interaction will be utilized using the physical browsing phenomenon. The main purpose of interacting via mobile device is to improve the security of transactions as well as eliminating the need to use cash/credit cards.

Fig. 1: Existing banking transactions system

3. ATM TECHNOLOGIES, STANDARDS AND SERVICES

ATM technologies, standards, and services are being applied in a wide range of networking environments, as described briefly below (Fig. 2)

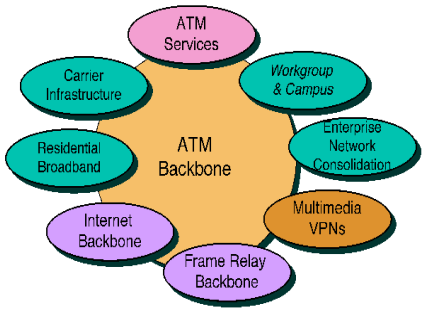


Fig.2: ATM Technologies, Standards and Services

- ATM services— Service providers globally are introducing or already offering ATM services to their business users.
- ATM workgroup and campus networks— Enterprise users are deploying ATM campus networks based on the ATM LANE standards. Workgroup ATM is more of a niche market with the wide acceptance of switched-Ethernet desktop technologies.

- ATM enterprise network consolidation — A new class of product has evolved as an ATM multimedia network-consolidation vehicle. It is called an ATM enterprise network switch (ENS). A full-featured ATM ENS offers a broad range of in-building (e.g., voice, video, LAN, and ATM) and wide-area interfaces (e.g., leased line, circuit switched, frame relay, and ATM at narrowband and broadband speeds) and supports ATM switching, voice networking, frame-relay SVCs, and integrated multiprotocol routing.

- Multimedia virtual private networks and managed services — Service providers are building on their ATM networks to offer a broad range of services. Examples include managed ATM, LAN, voice and video services (these being provided on a per-application basis, typically including customer-located equipment and offered on an end-to-end basis), and full-service virtual private-networking capabilities (these including integrated multimedia access and network management).

- Frame-relay backbones — Frame-relay service providers are deploying ATM backbones to meet the rapid growth of their frame-relay services to use as a networking infrastructure for a range of data services and to enable frame relay to ATM service interworking services.

- Internet backbones — Internet service providers are likewise deploying ATM backbones to meet the rapid growth of their frame-relay services, to use as a networking infrastructure for a range of data services, and to enable Internet class-of-service offerings and virtual private intranet services.

- Residential broadband networks — ATM is the networking infrastructure of choice for carriers establishing residential broadband services, driven by the need for highly scalable solutions.

- Carrier infrastructures for the telephone and private-line networks — Some carriers have identified opportunities to make more-effective use of their SONET/SDH fiber infrastructures by building an ATM infrastructure to carry their telephony and private-line traffic.

4. PROPOSED ATM BANKING SYSTEM

Biometric authentication has become more and more popular in the banking and finance sector. The idea of fingerprint is not only for security but also to overcome the lack of customer understanding on ATM concept. We proposed ATM with face recognition and fingerprint security system, in order to meet its customer's needs who many of them have savings account and need to have access to their money during non-banking hours. Operated using only a SIM card and a fingerprint scanner, the machines offer excellent security to card holders since there is very low possibility of fraud. If a customer loses the card, it is difficult for another person to use it because of the digital fingerprint and image of the original user. By using fingerprint recognition customers are more comfortable with the idea of saving their money with the bank because they understand that if they lose their ATM card, no one can replicate their fingerprint and take their money. Fingerprint authentication is the most popular method among biometric authentication, fingerprint based identification is one of the most mature and proven technique. In banking system Biometrics holds the promise of fast, easy-to-use, accurate, reliable, and less expensive authentication for a variety of applications. At the time of transaction customers enrolment their fingerprint to a high resolution fingerprint scanner. The fingerprint image is transmitted to the central server via secured channel. At the banking terminal the minutiae extraction and matching are performed to verify the presented fingerprint and image belongs to the claimed user in bank database. The authentication is signed if the minutiae matching are successful. The proposed scheme is fast and more secure. A basic biometric authentication system consists of five main components. These are: sensor, feature extractor, fingerprint/template database, and matcher and decision module. The function of the sensor is to scan the biometric trait of the user. The function of the feature extraction module is to extract the feature set from the scanned biometric trait. This feature set is then stored into the template database. The matcher modules takes two inputs, i.e. feature set from the template database and feature set of the user who wants to authenticate him and compares the similarity between the two sets. The last module, i.e., the verification module makes the decision about the matching of the two feature sets. Biometrics is a rapidly evolving technology that is being widely used in forensics, such as criminal identification and prison security, and that has the potential to be used in a large range of civilian application areas. Biometrics can be used to prevent unauthorized access to ATMs, cellular phones, smart cards, desktop PCs, workstations, and computer networks.

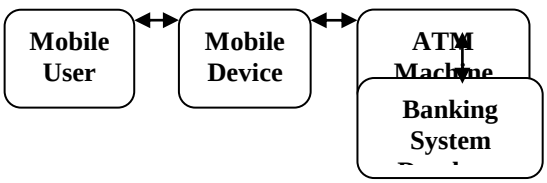


Fig. 3: Proposed ATM system

5.ARCHITECTURE

The overall architecture of this third generation ATM transaction using SIM card is described below:

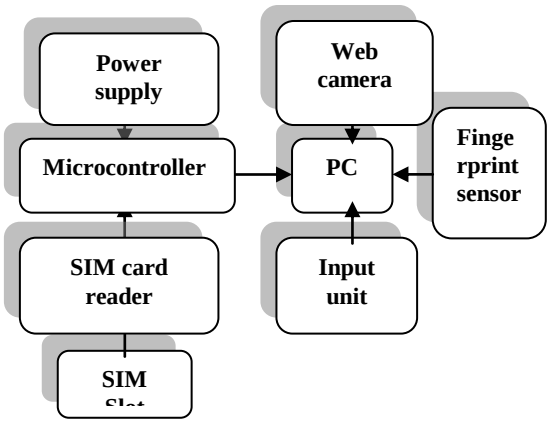


Fig.4:Architecture of third generation ATM

User's SIM card is inserted into the SIM slot unit. SIM card reader device scans SIM number through GSM Modem. GSM Modem is a data oriented GSM transceiver system that uses a network provider to connect and transfer data. GSM modem has on board RS-232 converter that you can use to connect the modem directly to a PC serial port. Using a windows accessory program HyperTerminal, you can immediately begin test driving your GSM Modem. The GSM Modem will work with just about any popular microcontroller.

Microcontroller with a built-in UART; this will save you from having to write bit banging codes that only for advanced programmers to consider. If your microcontroller works on 3.3V supply,
interfacing will be a breeze. Interfacing to a 5V microcontroller will require some
more work. The RX pin of the GSM Module is designed to accept 3.3V logic level only, forcing a 5V logic input may damage the GSM Modem. To
prevent this, a resistive divider must be added on the TX pin of your 5V microcontroller circuit. The GSM SIM card reader unit scans the mobile number, Which data given to the Teller machine for further processing. At the same time, web camera captures the images and comparing with the existing user's image using digital signal processing. And further it requires fingerprint for the access of money from ATM machine. Atmel Company's AT77CI04B be used as a fingerprint recognition. It has a 500dpi resolution, anti-press, anti-static, anti-corrosion. If an image, finger print & PIN number are same then further processing continued. Otherwise the process gets terminated.

6. SECURITY CONSIDERATION

The Secure M-Application deals with information transfer and financial transactions. Hence, the security elements considered during the design and implementation stage consist of making sure that the RFID and NFC used are in compliance with the ISO 14443 standards. WLAN technologies are equipped with the latest encryption protocol standard. WEP and WPA address these issues by providing password protection for access control and encryption for privacy. The wireless security protocol 802.11w which is to be introduced in April 2008 has promised to provide facilities that will prevent denial of service attacks and make use of the AES encryption standards. These protocols (WEP and WPA), only deal with access control and privacy issues. That is determining who is allowed to enter your network and hiding information from hackers who may try to intercept information during transmission. However because the Secure M-application is more susceptible to security threats, the design incorporates other security measures such as making use of Biometric data, as a further authentication layer. It can utilize any of the fourteen different types of biometrics that fall within two categories namely: those that measure behaviour and those that measure physical traits. Any of the Biometric information within these types can be used in identifying users, by making use of individual anatomy or physiology, that is either deeply ingrained into the skin, or other behavioural characteristics, or it can be a combination of the two. Hence, the data will be used as a unique personal attribute for security and authentication purposes. Since the focus of this paper is not on analyzing biometrics as a means of authentication, details on the use of biometric can be found in and using biometric data to generated encryption and decryption keys. Both authentication factors are needed to match the stored once before allowing access to transactions else authentication will be automatically denied. This process will be able to reduce the effects of brute force attack, as millions of combinations need to be tried before gaining access and as only 3 tries are allowed this should prevent such attacks from occurring. Also it is important to point out that a single sign-on process will improve the reliability of identity management and access control. Having this application as the security measure, the Secure M-Cash Withdrawal will be able to provide an excellent wireless identity management that will remove the current risk of identity theft and meet the required security standard of implementing the secure M-cash withdrawal and other M-applications.

7. USE OF ATM:

The following uses can be made by using both ATM card as well as SIM card

- **Cash Withdrawal and Balance Enquiry**
  In spite of a number of innovative services being made available at many ATMs, cash withdrawal stills remains the most accessed service at ATMs. However, the migration of routine bank transactions like cash withdrawals and balance enquiries from teller counters to ATMs significantly raises the potential for savings in employee costs and greater employee focus on value-added revenue-enhancing activities such as selling other financial products and advisory services to customers.

- **Cash /Cheque Deposit**
  Again, due to the strong cash culture in India, cash deposits are most likely higher than in other markets, especially cash deposits made by commercial

customers such as retail shopkeepers and those whose work involves substantial travelling. A high cash withdrawal rate results in higher ATM servicing costs due to frequent cash replenishment requirements. Recent developments in ATM technology have made it possible to recycle cash in ATMs. Currency notes received as cash deposits are counted; soiled notes separated and deposited cash dispensed to fulfil withdrawal transactions. However, regulatory concerns relating to identification of counterfeit notes and its depositors need to be addressed first. ATM with Cheque deposit facility is not picking up in India, like other countries. One of the reasons is the delay in collection of the cheque deposited in ATMs. Cheque deposited in ATMs is to be collected and deposited in the designated branch for collection. Another reason is the introduction of cheque deposit Kiosks by various Banks especially Private sector ones. These are kept at each some important locations/branches where customers can deposit their cheques which are collected at intervals which may be difficult in ATMs.

- **Bill Payments**

  Most utilities have inadequate infrastructure for receiving bill payments resulting in long queues at collection centres. Hence, bill payment at ATMs has achieved noticeable acceptance by bank customers. Most banks provide this service through bi-lateral arrangements with bill-payment service providers. ATM users register their water, electricity and telephone utility accounts with banks, check their dues at ATMs, approve bill payments that are debited to their bank accounts and receive printed receipts for the transactions. This service has the effect of improving customer satisfaction for both the bank as well as the bill-payment service providers. Some Banks' ATMs even accept charitable contributions to Temples.

- **Sale of Paper Based Products**

  ATMs are ideally suited to sell paper-based products and services such as tickets, wireless phone recharge cards, financial products, etc. The screen interface allows browsing and customization, access to bank accounts facilitate payments and printing capabilities produce the actual product/service. A number of banks including ICICI Bank, SBI and PNB have ATMs at Mumbai's local railway stations to dispense season tickets to commuters. Own-bank customers pay no extra charge while other bank customers pay a fee of Rs. 50 for this extremely useful service of anytime ticket purchase. Railway season tickets represent a high-volume mass-appeal product. As technical standards get established and product/service sellers become aware of the ATM sales channel, niche-appeal high-margin products like entertainment tickets will join the fray.

8. CONCLUSION

This paper presents the new idea that can be used as a means of interaction between mobile phone, ATM machine and a Banking application for the purpose of withdrawing cash. The proposed design ; the secure M-cash withdrawal allows the use of mobile phones as a tool of interaction and provide flexibility through a robust identity management architecture. The Secure M-cash has examined the possibility of making use of similar approaches/techniques (RFID and NFC) for other applications and already there are some applications that have adapted this strategy. The Secure M-Cash Withdrawal has been defined, it will form as a foundation for future work within this area, which includes implementing a PC based simulation of the architecture and implementing the system. Automatic Teller Machines have become a mature technology which provides financial services to an increasing segment of the population in many countries. Biometrics, and in particular fingerprint scanning, continues to gain acceptance as a reliable form of securing access through identification and verification processes. This paper identifies a high level model for the modification of existing ATM systems using both security protocols as PIN, Biometric fingerprint strategy & face recognition module. We have been able to develop a fingerprint mechanism as a biometric measure to enhance the security features of the ATM for effective banking transaction for E-banking system. The prototype of the developed application has been found promising on the account of its sensitivity to the recognition of the customer's finger print as contained in the database. This system when fully deployed will definitely reduce the rate of fraudulent activities on the ATM machines such that only the registered owner of a card access to the bank account.

REFERENCES

[1] Välkkynen, P. Korhonen, I., Plomp, J., Tuomisto, T., Cluitmans, L., Ailisto, H. and Seppä, H. "A user interaction paradigm for physical browsing and near-object control based on tags": in *Proc. Physical Interaction Workshop on Real World User Interfaces*, (2003), 31-34. Authorized licensed use limited to: Bharat University. Downloaded on October 10, 2009 at 08:53 from IEEE Xplore. Restrictions apply.

[2] *Lauri Pohjanheimo, Heikki Keränen and Heikki Ailisto,* "Implementing TouchMe Paradigm with a Mobile Phone", ACM International Conference Proceeding Series; Vol. 121.

[3] Keränen, H., Pohjanheimo, L., Ailisto, H. (2005) Tag Manager: a Mobile Phone Platform for Physical Selection Services in International conference on Pervasive Services (ICPS 2005). Santorini, Greece. pp 405-412.

[4] Synchronized Multimedia Integration Language (SMIL 2.0), W3C Recommendation 07 August 2001 http://www.w3.org/TR/smil20/ .

[5] Java 2 Platform, Micro Edition (J2ME). http://java.sun.com/j2me/ .

[6] M. Johns and G. Marsden *"Mobile Interaction Design"*, Wiley and Sons Ltd 2006, pg s178-182.

[7] "Openwave *Phone Simulator"* http://developer.openwave.com/dvl/tools_and_sdk/phone_simulator/ .

[8] European ATM Security Team, *"European ATM Security Team"*, Diebold, Incorporated 2006, http://www.diebold.com/rd/whitepapers/atmfraud&security.pdf

[9] A. Jain, R. Bolle, and S. Pankati, "*Biometrics: personal identification in networked society*", Kluwer Academic Publishers, 1998.

[10] J. Chirillo and S. Blaul, "*Implementing Biometric Security*", Wiley & Sons, 2003.

[11] S. Hoque, M.C. Fairhurst, F. Deravi, W.G.J. Howells, "*On the Feasibility of Generating Biometric Encryption Keys"* IEE Electronics Letters, 41(6), 309-311, 2005.

[12] Cellan-Jones, R., *London starts digital cash trial* in Technology correspondent, BBC News Online, http://news.bbc.co.uk/1/hi/technology/7117213.stm.2007: UK.

[13] Blythe, P.T. *Improving public transport ticketing through smart cards*. in Proceedings of the Institution of Civil Engineers: Municipal Engineer. 2004.

[14] New Barclaycard is touch-and-pay in BBC News,http://news.bbc.co.uk/go/pr/fr/-/1/hi/business/6945991.stm.2007.

Cloud Latency: A Cloud Based Latency Checking System for the Developers

Arun Kumar P L, Mr.Arun Kumar D R
M.Tech Student, Assistant Professor
Department of Computer Science and Engineering, VTU, Bengaluru-562157
arunkumarpl124@gmail.com,arunkumardr1987@gmail.com

ABSTRACT: Now a days, the cloud platform plays a very important roles in development environment to implement various applications over servers. The developer should to know the latency to select appropriate server which in turn takes less time and perform the task very quickly. The main objective of this paper is to help the developers to check the latency of Amazon web services (AWS). By knowing the latency of perfect web server it helps the developers in selecting the particular server for their development of applications on cloud based system. It also helps the developers to optimize their services and solutions. Amazon has web server in different countries So, the latency also vary based on various locations of the Amazon servers. In the proposed system we finding the latency of servers based on the parameters such as region name, location of the server and endpoint of the service. Hence, the low latency is one of the main performance parameter for performance optimization in the cloud environment for the developers.

Keywords: Latency, AWS (Amazon web servers), optimization, performance, developers.

1. INTRODUCTION

Latency is the delay is measured from giving input into a system to get the desired outcome. The term is understood slightly different in various contexts and latency issues also vary from one system to another [1]. Latency greatly affects how usable and enjoyable electronic and mechanical devices as well as communications are. Latency in communication is demonstrated in live transmissions from various points on the earth as the communication hops between a ground transmitter and a satellite and from a satellite to a receiver each take time [2]. People connecting from distances to these live events can be seen to have to wait for responses. This latency is the wait time introduced by the signal travelling the geographical distance as well as over the various pieces of communications equipment [3].

Network latency is an expression of how much time it takes for a packet of data to get from one designated point to another. In some environments (for example, AT&T), latency is measured by sending a packet that is returned to the sender; the round-trip time is considered the latency. Ideally latency is as close to zero as possible [4].

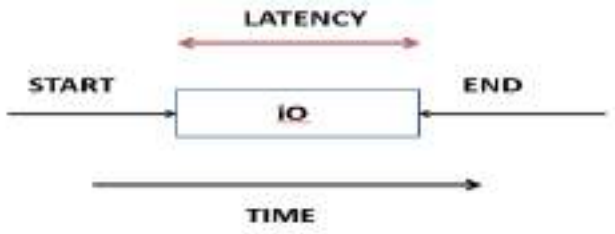


Fig1: Identitical way of representation of latency over a time.

AWS is located in 12 geographical regions. US East (North virginia), where the majority of AWS servers are based, US West (northern california), US West(Oregon), Southeast Asia (singapore), East Asia (Tokyo, seoul, Beijing) and

Australia (Sydney). There is also a "GovCloud", based in the Northwestern United States, provided for U.S. government customers, complementing existing government agencies already using the US east region. Each region wholly contained within a single country and all of its data and services stay within the designated region. Each region has multiple 'Availability zones', which are distinct data centers providing AWS services [15]. Availability zones are isolated from each other to prevent outage from spreading between zones. Several services operate across availability zones while other can be configure to replicate across zones to spread demand and avoid down time from the failures [7]. As in the year 2014, Amazon web services operated an estimated 1.4 million server across 28 availability zones [6].

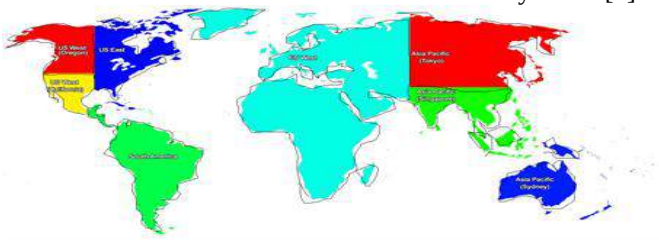


Fig2: Amazon web servers region map

2. EXISTING SYSTEM

Since, the Existing systems are desktop based. We can check the latency of the server using command prompt. But the same is not for the developers as a service. And finding the latency is not based on the cloud and latency is not available in the mobility state. Its only supports static type of allowance of network performance. It is limited for checking only the latency of the servers which is much more important server feature based over location [8]. This System has immense future development scope as it can be integrated with more subsystems for checking cloud features which are still missing. Due to unknowing of latency of cloud many of the organization scale rapidly decreasing the profit based on latency of the cloud [10].

3. PROCESS DESCRIPTION

Using Amazon web server locations and their online identity java-script will be done by giving a ping commands to those

servers and their results will be automatically captured as values into tables. Administrator of the system can select for Amazon EC2, simple database and other services in the system and can check for their latency over different servers located in different locations over globe [11]. By selecting low latency of perfect server it helps to improve the optimization and to choose server region.

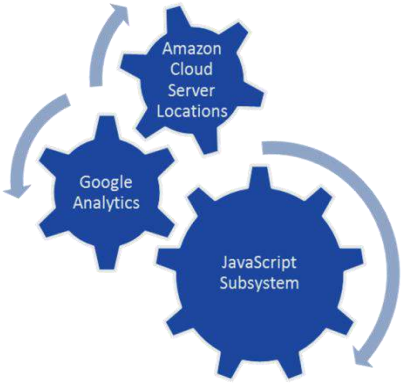


Fig3: Cloud server location analysis

The figure 3 shows the analysis of the latency by making use of Google analytics and Amazon server location name. Google analytics analyze the region name, region location, and end point of the regions. it helps to locate exact server of Amazon over a globe [12].

4. AMAZON WEB SERVICES

Amazon web server is a collection of cloud computing services that make up the on-demand computing platform offered by amazon.com. AWS provides some services based on cloud platform. They are Elastic Cloud Computing (EC2), Simple database, simple storage service (S3), Simple Query Service (SQS), Simple Notification Service (SNS) [15].

a. Elastic cloud computing(EC2)

Amazon Elastic Compute Cloud (EC2) forms a central part of Amazon.com's cloud-computing platform, Amazon Web Services (AWS), by allowing users to rent virtual computers on which to run their own computer applications. EC2 encourages scalable deployment of applications by providing a web service through which a user can boot an Amazon Machine Image to configure a virtual machine, which Amazon calls an "instance", containing any software desired [7]. A user can create, launch, and terminate server-instances as needed, paying by the hour for active servers - hence the term "elastic". EC2 provides users with control over the geographical location of instances that allows for latency optimization and high levels of redundancy [13][14].

b. Amazon s3

Amazon Simple Storage Service (Amazon S3), provides developers and IT teams with secure, durable, highly-scalable cloud storage. Amazon S3 is easy to use object storage, with a simple web service interface to store and retrieve any amount of data from anywhere on the web. With Amazon S3, you pay only for the storage you actually use. There is no minimum fee and no setup cost.

Amazon S3 offers a range of storage classes designed for different use cases including Amazon S3 Standard for general-purpose storage of frequently accessed data, Amazon S3 Standard - Infrequent Access (Standard - IA) for long-lived, but less frequently accessed data, and Amazon Glacier for long-term archive. Amazon S3 also offers configurable lifecycle policies for managing your data throughout its lifecycle. Once a policy is set, your data will automatically migrate to the most appropriate storage class without any changes to your applications.

Amazon S3 can be used alone or together with other AWS services such as Amazon Elastic Compute Cloud (Amazon EC2) and AWS Identity and Access Management (IAM), as well as data migration services and gateways for initial or ongoing data ingestion. Amazon S3 provides cost-effective object storage for a wide variety of use cases including backup and recovery, near line archive, big data analytics, disaster recovery, cloud applications, and content distribution,

c. Simple Notification Service(SNS)

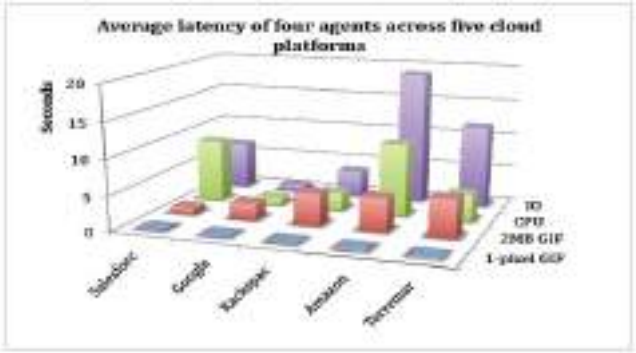
Amazon Simple DB is a distributed database by Amazon.com. It is used as a web service in concert with Amazon Elastic Compute Cloud (EC2) and Amazon S3 and is part of Amazon Web Services as with EC2 and S3, Amazon charges fees for Simple DB storage, transfer, and throughput over the Internet.

5. Proposed System

A cloud based system for checking latency of Amazon cloud servers based on different locations is just what the developer are in need of for the performance optimization. It will help them choose the exact server as required. There are many systems for the developers but the system for checking latency is still missing. The implementation using simple resources from Google analytics and Amazon server location details can make it possible.

Cloud platforms are a cloud computing platform by Salesforce, google, amazon, rackspace etc, . These platform offers hosting on same supporting infrastructure that these uses internally for end user products. Cloud Platform provides developer products[3] to build a range of programs from simple websites to complex applications.

The graph1 shows that the average latency of cloud platforms. The latency is detected by speed of the network performance over the time. Compare to other cloud platform amazon is the best cloud platform due to the less delay for different operation performed in between source and destination point [17,18]. which is seen by a user.



Graph1: Comparision of latency of different cloud platforms.

6. IMPLEMENTATION

Checking the latency of the different server is make it possible by google analytics and java script sub system. Google analytics analyses the location of the server and the region name. This system will have massive future if we are integrated with subsystem for checking the cloud features and details of the server located regions.The Figure 5 shows the collaboration of MVC components. Cloud latency has mainly three components they are user, system, and latency. These have a various sub components which are responsible for analysis of delay of the servers. User undergoes to select different services of the amazon web server (AWS). System has front and and back-end view.

That's javascript, html, css. Latency is analysed by a location of the service and the location value measured by timestamp. The analysis performed through a system.

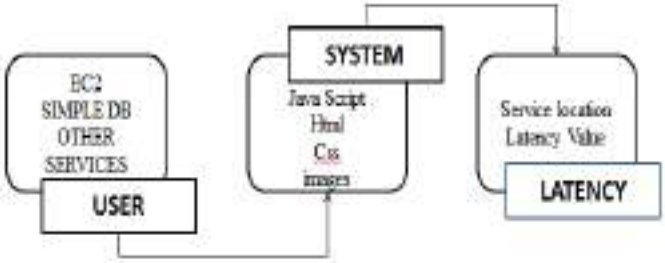


Fig 4: Analysis view of cloud latency

6.1 Methodology for the architecture

Model view controller is a software architectural pattern mostly for implementing user interfaces on computers. It divides a given software application into three interconnecred parts, so as to separate internal representations of information from the ways that information is presented to or accepted from the user.

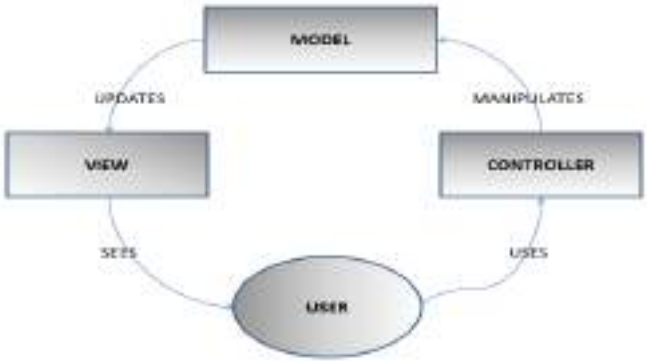


Fig 5: Collaboration of MVC components.

Figure 5 expresses the "core of the solution" to a problem while allowing it to be adapted for each system. MVC architecture can vary significantly from traditional description.

The MVC are differentiating as, model as cloud server and view as browser and controller is admin in our system. User making the use of controller to manipulates the model at each system level. Model updates the view that's sees by the user. The model works in the loop fashion.

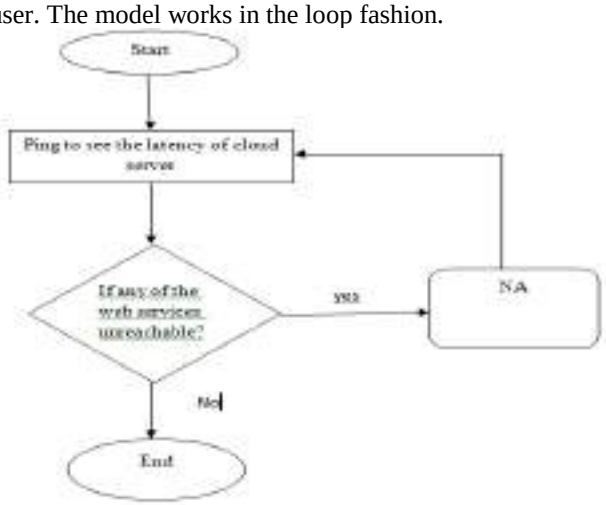


Fig 6: Flow process of cloud latency

Above figure 6 shows the latency process from end- end scenario. Ping is a command to get a latency of web server. The work process of cloud latency tends to get a latency time of each server. The phenomenon runs based on checking the latency of servers in different regions of AWS. If the service available the server will give a timestamp from source and destination. If the service is unreachable to destination point, then it will show NA. then the process again goes to requesting phase with different services. The process will continues upto get a desired latency from services provide by Amazon.

6.1 RESULT

Cloud latency helps for the developers to check the latency of AWS servers. Latency helps to choose the accurate server which takes minimum latency based on data transmission between end points. Resulting outcome explains the choosing servers of Amazon for the different operations based on different services.

7. CONCLUSION

A cloud based system for checking latency of Amazon cloud servers based on different locations is just what developer are in need of for performance optimization. It will help them to choose the exact server as they required. There are many systems in cloud for developers but, system for checking latency is still missing. Implementation using simple resources from Google analytics and Amazon server location detail can make it possible. The System will choose the exact server based on services providing by Amazon web server which is located over a globe.

ACKNOWLEDGEMENT

We gracefully thank our college SVCE for providing us with all the necessary help and grooming up in to be Master of Technology. I express my sincere gratitude to Dr, C. Prabhakar Reddy, Principal, SVCE and Dr. Suresha, HOD, Dept. of CSE, SVCE, Bangalore forproviding the required facility and giving me an opportunity to work on this topic. I also extend my sincere thanks to my guide Mr. Arun kumar D R, Asst. Professor, Dept. of CSE, SVCE for support and guiding me to work on this topic

Traffic Profiling and Dynamic weight computing using dWRR for Average packet delay versus traffic load

[1]Akshatha Kamath, [2]Humera Siddiqua, [3]Usha M. S
Asst. Professor, Sri Krishna Institute of Technology
kamathakshu@gmail.com, humera.siddiqua928@gmail.com, usha.malligere@gmail.com

Abstract:- The capacity of the conventional wireless mesh network (WMN) with single channel single radio is limited due to co-channel interference. To resolve this capacity limitation problem, multi-channel multi-radio (MCMR) protocols have been proposed. Wireless mesh networks has attractive significant attention towards promising technology. It is becoming a trend for the fourth generation of wireless mobility. Communication in large-scale wireless networks can create bottlenecks for scalable implementations of computationally intensive applications. Classes of crucially important communication patterns that have already received considerable attention in this regard are group communication operations, since these inevitably place a high demand on network bandwidth and have a consequent impact on algorithm execution times. Multicast communication has been among the most primitive group capabilities of any message passing networks. Existing solutions offered for providing multicast communications in WMN have severe restriction in terms of almost all performance characteristics. Consequently, there is a need for the design and analysis of new efficient multicast communication schemes for this promising network technology. An IEEE 802.11s standard draft has been lately designed to define the WMNs framework and architecture. Multi-channel multi-radio wireless mesh networks (MCMR WMNs) have been introduced to improve system performance. However, interfaces and channels management in MCMR WMNs becomes essential to achieve the network performances rise. In this paper, we propose a solution that defines a MAC layer module named Interface Management Module (IMM) under a hybrid channel assignment. The IMM manage the multi-channel radio interface based on scheduling algorithms. A dynamic scheduling algorithm has been proposed. The algorithm is based on traffic profiling and dynamic weights computing for channels scheduling.

Keywords- multicast; Wireless Mesh Network; Interface Management Module; Traffic Profiling

I. INTRODUCTION

Currently accessing the internet, working on internet all over the world has been increased, especially in the developed and developing countries. Wireless Mesh Networks (WMNs) have emerged as a advanced technology as a backbone for a wide range of applications. Their main feature is autonomy. They can be formed without the need of any infrastructure suitable for using in scenarios such as communication in natural disaster (infrastructure damage), home networking, etc.

Now a days, the demand for group communication technology has significantly increased. More and more people prefer to watch football matches and TV drama from the internet rather than from traditional TV. As a technology of group communication, the aim of multicast is to send information from the source sender to multiple receivers. Therefore, multicast is used by the service provider to deliver service to multiple subscribers.

On the other side, the popularity of wireless local area networks (WLANs) based on IEEE 802.11 has been significantly increased [1]. Consequently, the interest over networks based on this topology has also been raised. Thus, IEEE 802.11-based WMNs are being actively explored. Thus, an IEEE 802.11 task group, i.e. 802.11s, was formed. Its aim and essential motivation is to specify a flexible and extensible standard for WMNs based on IEEE 802.11 and to provide solutions by which the WMN backbone could be built without the need of any centralized administration and with minimal configuration difficulties. The last released draft defines the MAC and PHY layer specifications of the 802.11 based wireless mesh networks [2]. It presents WMN framework architecture. A WMN is composed of wireless mesh nodes, commonly known as Mesh Points (MPs) which form a self-contained network with wireless backbone access: Mesh Basic Service Set (MBSS). Mesh stations communicate with each other through single-hop or multi-hop paths in a peer-to-peer fashion. They are either stationary or with low mobility and have ample energy supply. A MP may be collocated with one or more other entities. It is a Mesh Portal (MPP) while it is configured as a gateway allowing access to external networks such as wired LANs. MP can also handle access point functionalities, i.e. Mesh Access Points (MAP), while associated with wireless network interface card. Besides, WMNs can be deployed in a multi-channel multi-radio fashion; commonly known as multi-channel multi-radio wireless mesh networks (MCMR WMNs). MCMR WMNs support MCMR mesh nodes where MPs are equipped with multiple radio interfaces, interfaces can be tuned on a different frequency channel. A Unified Channel Graph (UCG) is a set of MPs that have links via a common wireless medium communication. According to [2], a multi-channel mesh network is composed of several UCGs and a single channel MBSS has only one UCG. Moreover, a multi-radio mesh node belongs to different UCGs because the same device is composed of multiple radio interfaces, each one is tuned on a specific UCG. A mesh node may change its operating channel

and switch from one UCG to another, using the Channel Switch Protocol (CSP) [2]. Channel switching causes are numerous (e.g. high level of interferences, etc).

Description about multi-channel, multi-radio multicast requirement this paper. Section II examines the proposed solutions for interfaces and channels managements in a MCMR WMN; followed by section III which describes implementation of proposed method. Section IV presents simulation results while implementing our solution. Finally, the last section concludes this paper.

II. RELATED WORKS

Some methods have been proposed to resolve the interfaces and channels management problem in MCMR WMNs (i.e. which interface(s) and channel(s) to use for traffic exchanging with neighboring nodes). Most methods require having as many interfaces as the number of channels for each node which prove to be expensive in the real system implementation. Few methods propose solutions for nodes with interfaces fewer than channels.

The challenges posed by the continuously growing need for delivering efficient multicast communication over wmn .this study presents a new load balancing aware multicast algorithm with the aim of enhancing the QOS in the multicast communication over WMN's . In this paper a Gateway-cluster based Load Balancing Multicast algorithm(GLBM) investigate the load balancing problem of multicast in the WMNs is proposed. The multicast applications such as multicast conference & multicast TV require instant real time communication & large packet size. Due to these concerns priority metrics are end to end delay and throughput. Algorithm focuses on high throughput and low end to end delay multicast session through achieving load balancing[1].

In this paper[2] ,introduction of new voice call capacity model of hybrid multi-channel protocol(HMCP) on Multi-Channel Multi-Radio(MCMR) WMNs .Both experimental and simulation results demonstrate that the proposed call capacity model accurately estimates the voice call capacity for G.711 & G.729 codes. Propose 2 QoS routing algorithm for finding feasible routes to meet QoS constraints as well as to improve the call capacity of network by utilizing the proposed call capacity model. Proposed QOS routing algorithms effectively protect voice calls and increase the call capacity[2].

In this paper [3], we propose a solution that defines a MAC layer module named Interface Management Module (IMM) under a hybrid channel assignment. The IMM manage the multi-channel radio interface based on scheduling algorithms. A dynamic scheduling algorithm has been proposed. The algorithm is based on traffic profiling and dynamic weights computing for channels scheduling.

III IMPLEMENTATION

In this section, we address the interfaces and channels management dilemma in a MCMR environment.

using a hybrid channel assignment strategy where the first interface is a fixed interface with a static channel assignment; while the second interface is a switchable interface which uses the dynamic channel assignment strategy [9]. Thanks to this interfaces' assignment approach full duplex mode can be achieved. In fact, the MCMR mesh node fixed interface is primarily used for receiving data from neighbors while its dynamic interface is dedicated to transmit data to its neighboring nodes. The switchable interface is tuned on a channel which may be changed at any time. Thus, if two mesh nodes need to communicate for exchanging data, the switchable interface of the sender node and the fixed interface of the receiver node must be tuned on the same channel. If it's not, the sender node switchable interface switches on the channel on which the receiver node fixed interface is tuned. Radio interfaces coordination for channel switching is handled by CSP. This work focuses on defining channels management module for the switchable interface for decision coordination of which and when the interface has to be tuned on that channel.

Switching the radio interface from one channel to another incurs a non-negligible delay. Consequently, too frequent channel-switching may significantly degrade network performances. According to [10], the channel-switching delay varies from 200 µs to 20 ms. Therefore, when an interface is not tuned on the correct channel, packets have to be buffered in

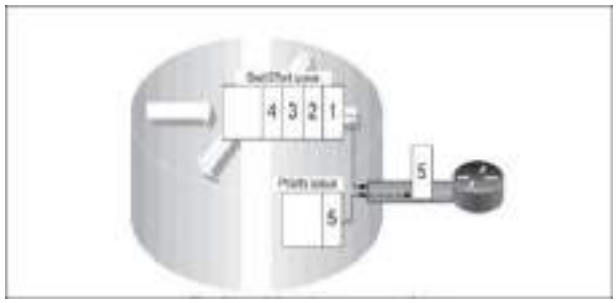


Fig1.Priority and profile mapping

a channel queue until the interface switches on the corresponding channel. Thus, there is a need to use a queuing algorithm to buffer packets, as well as a scheduling algorithm to transmit these buffered packets. The scheduling algorithm decides on which channel does the switchable interface has to be tuned on and for how long. It must also reduce the frequent switching and ensure that the queuing delay is not very large. Therefore, due to the non-negligible channel switching delay, introduces the *T*min parameter. *T*min is the minimum channel scheduling time[3]. The aim of the traffic profiling is to profile the traffic received and buffered in the channels' queues. The profile characterizes the packet priority. Three traffic profiles are considered: Voice, Video and Best Effort (BE) profiles. Traffic priority information is carried in the packet header. In fact, the frame format includes '*QoS Control Field*' which contains TID sub-field that specifies the packet priority [1]. The TID sub-field is set to a value between 0 and 7(table 1).

TID sub-field value is used then to profile traffics. A mapping between profile and packet priority is done . BE

A. System Model

Fig2.Module Management for interface

The 'Scheduler' computes the service time *TS* of each channel and selects the channel queue to service when the *TS* of the current channel ends. The network card of the switchable interface switches then on the selected channel and transmits its corresponding packets during its *TS* . The 'Scheduler' uses a scheduling algorithm to select the channel to serve. In this paper, we study two well -known and basic scheduling algorithms: round robin (RR) and weighted RR (WRR), here named as static WRR.

B. Scheduling algorithms

During the implementation of the interface module, there is one queue for every channel. Each queue buffers the packets that have to be sent on the corresponding channel. channel scheduling methods used are.

Static Weighted Round Robin

The Round Robin (RR) is a scheduling algorithm that assigns to active channels equal portions of service time in circular order without handling the queued packets priority. An active channel is a channel which has packets in its buffering queue. Thus, RR algorithm allows every active channel to take turn in transferring packets on the shared switchable interface in a periodically repeated order. If one channel queue is out of packets, the next channel queue takes its place. The RR algorithm provides fairness among all channels' queues, i.e. all channels have equal weights, so equal service times *TS* . The *TS* value is assigned by the 'Scheduler' and must be at least *T*min .

With Static Weighted Round Robin (sWRR) scheduling algorithm, different static weights are affected to each active. They are inflexible with traffic load variation and insensitive to queued packets priority. The channel service times are computed for each channel using its static weight and (1) and (2). Switchable interface tunes then on every channel during its *TS* . First, the 'Scheduler' stores*W*min :

$$W_{min} = \min_{1<i<N} \{Wi\} \quad (1)$$

where *W*min is the the channels' smallest weight, *Wi* is the weight of the channel i and *N* is the number of buffering queues. Then, it computes the service time of all the channels:

$$T_i = \frac{T_{min}}{W_{min}} * W_i \quad (2)$$

where TS_i is the service time of the channel i, T_{min} is the minimum service time allocated to the channels.

With the sWRR scheduling algorithm, channels haven't got equal weights nor service times. Moreover, *TS* values are static and don't consider the traffic flows variation. Therefore, they may be not adapted to the channels load in some traffic load configuration which causes network performances degradation.

Dynamic Weighted Round Robin

Dynamic Weighted Round Robin (dWRR) & EDCA scheduling algorithm jointly considers EDCA features for QoS-based services differentiation and channels weights dynamicity.

Weights Computing:

After profiling the traffics, the channels' weights are computed. To this end, coefficients are assigned to each traffic profile. These coefficients are then used to compute the dynamic channels weights by applying an additive metric. Higher priority profiles have got bigger coefficient ($\alpha < \beta < \gamma$), see Table I. Dynamic channel weights are then computed using (3) and (4):

$$Coef_i = \sum_{j=1}^{Ki} Coef_{ij} \quad (3)$$

where *Coefi* is the coefficient of the channel i, *Coefij* is the coefficient of the flow j in the channel i and *Ki* is the number of traffic flows buffered in queue i. The dynamic weights are then computed using this formula:

$$Wi = \frac{Coef_i}{\sum_{j=1}^{N} Coef_j} \quad (4)$$

where *Wi* is the dynamic weight of the channel i,coefficient of the channel i and *N* is the number of buffering queues. After computing the channels' dynamic weights, channels' service times are computed using (1) and (2). With traffic profiling and dynamic weights computing, channels' weights become adapted to buffered packets priority. Moreover, channels weights are dynamically updated while traffic load and priorities vary to achieve adaptability with incoming flows.[3]

The 'Scheduler' sub-module uses one of the three aforementioned scheduling methods to select the next

channel on which the dynamic interface will be tuned on. The dynamic interface serves channels for an amount of time that equals their corresponding service times. It uses then EDCA mechanism to attempt to access to the wireless shared medium.[3]

IV. SIMULATION RESULTS

A. *Network Topology*

To compare the performance of the aforementioned scheduling algorithms, the network topology depicted in Fig. 2 is used. The MBSS is composed of 4 static mesh nodes. It is a simple topology that is designed to avoid routing problems and bad links troubles which are caused by a large network topology. A mesh node has two radio interfaces configured to use EDCA medium access mechanism. The radio band is 2.4 GHz. Thus, there are 3 orthogonal frequency channels. According to [13], in real case, the number orthogonal channels dependents on the hardware chosen and the distance between radios. Consequently, in the simulations, we place mesh nodes far enough distant that inter-channel interferences are neglected.

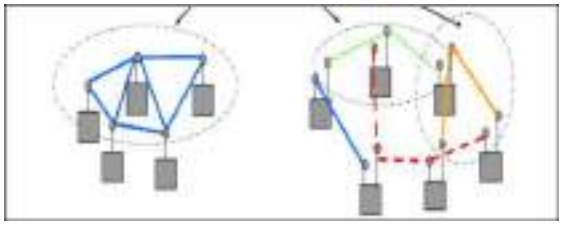


Figure 3. Network Topology

Every node is equipped with two radio interfaces using a hybrid assignment strategy. The static interface, dedicated to traffic reception, is configured to one of three available orthogonal frequency channels while the dynamic interface, may switch and tune on any of these orthogonal channels.

Flows' source node is Node A (c.f. Fig. 4). For sack of simplicity, all flows are CBR traffics with different priority. Node A sends Voice (AC_VO) and Video (AC_VI), Voice and BE, Voice and BK traffics to Node B, C and D respectively. IMM is implemented in all nodes but simulation results are collected from node A as the MCMR interfaces management appears only in this node using this configuration.

B. *Simulations Results*

With sWRR scheduling algorithm, channels' weights are invariable with traffic load variation. Thus, channels' weights

might be maladjusted to queues load due to an unlucky weights assignment or, could be adapted in some other cases. To show the consequence of an unlucky weights choice maladjusted to the channels' traffic priority and load, two simulation scenarios are examined. The first scenario is an

dWRR reduces VoIP latency in comparison with the remaining scheduling algorithms because channels carrying high priority traffics have higher time services during the scheduling decision than those with low priority traffics.

Unlucky Weights Choice

(UWC) of sWRR channels' weights where weights are not adapted to the traffic priorities of buffered packets. The second scenario is a Lucky Weights Choice (LWC), i.e. channels' weights and packets priorities are proportional. These scenarios point out adequate channels' weights importance.

The dWRR channel scheduling algorithm adapts channels' weights to its channels' traffic priorities and load, see Fig. 6.

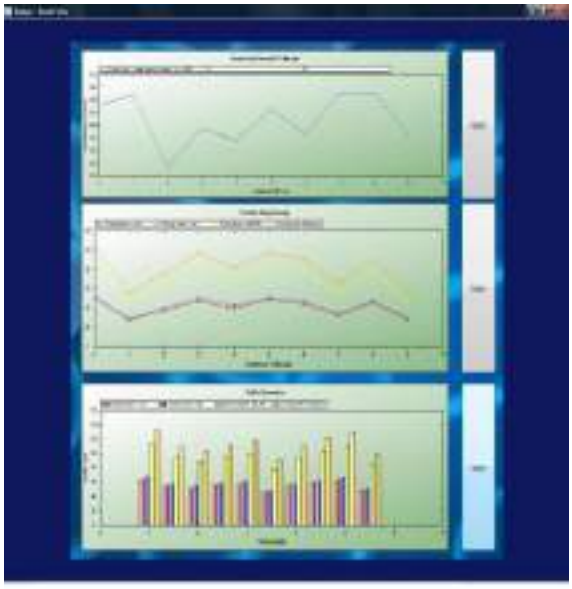


Fig4.Traffic Generation Graph for sRR

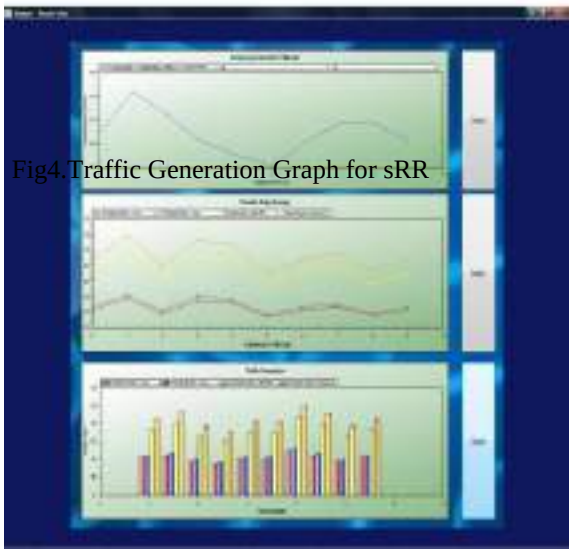


Fig 5.Traffic Generation Graph for dWRR

Moreover, while applying dWRR channel scheduling algorithm, the delay average of voice traffic is kept almost constant. In fact, when traffic load increases, dWRR adapts channels' weights to the traffic buffered in its queues. As a consequence, channels that handle delay sensitive traffics have priority to access to the shared resource. Therefore end to end delays decreases considerably while comparing to other scheduling scenarios. Besides, in sWRR and RR channel scheduling scenarios, delay average increases significantly as the network load rises and it exceeds 150ms (the voice delay constraint) which causes quality of service degradation. Delays increase is caused by constant weights that are kept invariant which may generate channels' time services maladjusted to traffic load and priority carried in channels' queues.

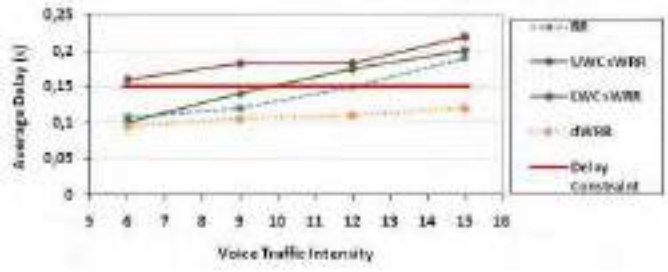


Fig6. Delay Average Vs Traffic Load

V. CONCULSION

In this paper, we have presented an interfaces and channels management method for multi-channel multi-radio wireless mesh networks. This method introduces an Interface Management Module (IMM) that uses scheduling algorithms to serve channels. Three scheduling algorithms have been studied. dWRR algorithm assigns dynamic channels' weights which are adapted to the traffic priorities and channels' load.

Simulations have shown that a bad weights choice affects considerably the network performances while using sWRR scheduling algorithm. dWRR presents the best simulation results because it affects adequate weights to the channels. As future works, dWRR can be further performed by affecting dynamic coefficients to traffic profiles.

REFERENCES

[1] A QOS aware multicast algorithm for wireless mesh networks Liang Zhao;Al-Dubai, A.Y:Geyong MinParallel & Distributed Processing. 2009 IPDS 2009.IEEE International Symposium on Year:2009 Pages:1-8,DOJ:10.1109/IPDPS.2009.5161193.

[2] Call capacity model and QoS routing algorithms on multi-channel multi-radio wireless meshes Sunghum Kim;Myungln Ji;Joong Sao Ma Communications(APCC),2012 18th Asia-Pacific Conference on Year:2012,Pages:311-316,DOI:10.1109/APCC.2012.6388153.

[3] A load and QoS aware scheduling algorithm for multi-channel multi-radiowireless mesh networks Zhioua G e m;Tabbane N, Wireless Communication and Networking Conference (WCNC),2012 IEEE Year:2012,Pages:2043-2047,DOI:10.1109/WCNC.2012.6214126.

[4] IEEE Standard for Information technology-Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11™, 12 june 2007

[5] IEEE P802.11s/D11.0,Draft Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks - Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Amendment 10: Mesh Networking, April 2011.

[6] Chaudhry, A.U.; Hafez, R.H.M.; Aboul-Magd, O.; Mahmoud, S.A.; "Throughput Improvement in Multi-Radio Multi-Channel 802.11a-Based Wireless Mesh Networks", GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference, 6-10 Dec 2010, pp. 1-5

[7] M. C. Castro, P. Dely, A. J. Kassler and N. H. Vaidya. "Qos-aware channel scheduling for multi-radio multi-channel wireless mesh networks", WiNTECH, ACM, 2009.

[8] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for ieee 802.11 wireless networks". Proceedings of IEEE Broadnets, October 2004, pp. 344-354,

[9] Franklin, A.A.; Balachandran, A.; Murthy, C.; Marina, M.; INFOCOM IEEE Conference on Computer Communications Workshops , 2010, March 2010, pp. 1-6

[10] S. Jagadeesan, B.S. Manoj and C. Siva Ram Murthy, ''Interleaved Carrier Sense Multiple Access: An Efficient MAC Protocol for Ad hoc Wireless Networks", Proceedings of IEEE ICC 2003, May 2003, pp. 1124–1128

[11] Uddin, M.F.; Alazemi, H.; Assi, C.; "Joint Routing, Scheduling and Variable-Width Channel allocation for Multi-hop WMNs", Communications (ICC), 2010 IEEE International Conference on, May 2010, pp.1-6

[12] J. MyungIn, K. Yonggyu, S. Myunghwan, and M. Joongsoo, "A novel multichannel multi-radiowireless mesh node architecture for ns-2", Proceedings of the 2009 WRI International Conference on Communications and Mobile Computing, IEEE Computer Society ACM, Vol. 2,

pp.147-151, 2009.

[13] A. Mishra, V. Shrivastava, D. Agrawal, S. Banerjee, and S. Ganguly, "Distributed channel management in uncoordinated wireless environments", In Proc. Of MobiCom 06, New York, USA, ACM, pp. 170-181, 2006.

[14] The Network Simulator ns-2, http://www.isi.edu/nsnam.

[15] R. A. Calvo and J. P. Campo, "Adding Multiple Interface Support in NS-2," University of Cantabria, Jan. 2007.

[16] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for ieee 802.11 wireless networks". Proceedings of IEEE Broadnets, October 2004, pp. 344-354,

[17] S. Ganguly, V. Navda, K. Kim, A. Kashyap, D. Niculescu, R. Izmailov, S. Hong and S. Das. "Performance optimizations for deploying VoIP services in mesh networks". IEEE Journal on Selected Areas in Communication JSAC, 2006

A Cloud-Based Venue Recommendation Framework on MobiContext

Ambika L G , Anitha T N,Bharathi M

PG scholar[1], Associate professor [2,3],Department of Computer Science and Engineering, S.J.C Institute of Technology, Chickballapur -562101.

ambikalg1991@gmail.com

ABSTRACT-In this paper demonstrate MobiContext, a hybrid cloud-based Bi-Objective Recommendation Framework(BORF).The MobiContext utilizes multi-objective optimization techniques to create modified recommendations.To address the issues pertaining to cold start and datasparse,the BORF performs data preprocessing by using the Hub-Average(HA) inference method and the Weighted Sum Approachis implemented for scalar optimization and vector optimization to provide finest suggestions to the users about a venue.The results of complete experiments on a large-scale real data set verify the accuracy of the proposed recommendation framework.

Index Terms—Multi-objective optimization, Collaborative Filtering(CF).

1. INTRODUCTION

In recent year, on going fast development of the Internet and simple accessibility of various e-commerce and informal organizations services,such as Amazon,Foursquare,and Gowalla,have sheer volume of information gathered by the administration suppliers on consistent schedule. The persistent gathering of huge volumes of information has moved the center of examination group from the fundamental data recovery issue to the separating of appropriate information[1], there fore, most research is currently coordinated towards the planning of more perceptive and self-sufficient data recovery frameworks, known as Recommendation Systems.

In late years,mobile social networking services, such as,google latitude,facebook has essentially picked up the fascination of countless.A portable informal community administrations permits a client to perform vast number of"check-ins" based on every day registration aggregate a huge volumes of data.Based on the information stored,several venue reccomendation system(VRS) were developed.So that VRS are intended to perform proposal of venues to users that most nearly match with users preferences.

A. PROBLEMS

The VRS are suffer with numerous limitations and challenges.A major challenges for such system is to process data at a real time and extract preferred venues from a huge dataset of user's historical checkins[3][1].The solution for VRS applied collaborative filtering(CF).The CF-based approache generate recommendation based on the similarity in actions and routines of users.

Later the CF-based recommendation system suffers from several limitations.The following are the most common problems in many existing CF-based recommendation systems:

B. CONTRIBUTIONS

To solve this problems we propose MobiContext,a hybrid cloud based Bi-objective recommendation framework(BORF).To address the cold start problem, it utilizing model-based Hub-average inference method,The HA method comptes and assignes popularity ranking to venues and users at various geographical location.

To address data sparseness,it utilize a metric known as confidence measure.confidence measures define the conditional probability.confidence measure is utilized to compute link waight between two users.This way,it helps replacing among zero similarity entries in user-to-user to matrix by alternate non-zero entries,to improving recommendation quality.

To improve scalability performance the cloud based MobiContext framework follows the Software as a Service(SaaS).The proposed framework can scale on demand as additional virtual machines are created and deployed.

2. SYSTEM OVERVIEW

The existing recommendation systems utilize centralized architectures that are not scalable enough to process large volume of geographically distributed data.

Therefore, to address the scalability issue, we introduce the cloud-based MobiContext BORF framework.

In terms of functionality,the proposed system architecture has main phases:a)A pre-processing phase b)A recommendation phase.Preprocessing phase is divided into a)Ranking module b)Mapping module.

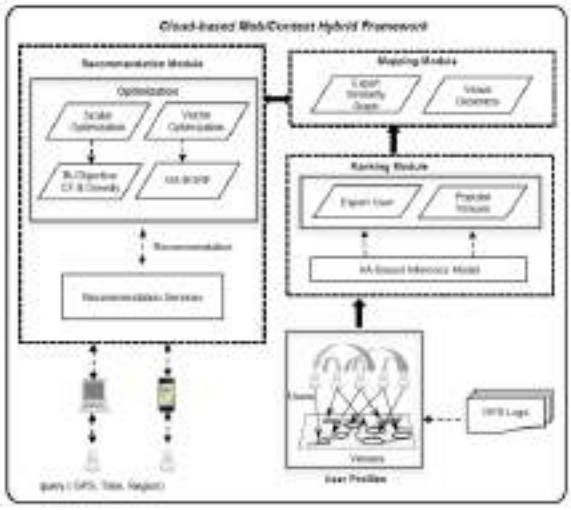


Fig 1:Cloud based MobiContext BORF framework

a)In ranking module the HA inference model is applied on user profiles to compute ranking for users and venue check-ins that is utilized to compute popularity ranking scores for users and venues.Which user visited more popular venues that user is called expert user.The all expert user visited venues that venue called popular venues.b)The mapping module computes the similarity among the expert users.It also computes the geographical distance of the current user from the popular venues.The user will gives query(GPS,location of user time and region)through mobile or desktop the recommendation service recives user queries.The recommendation service passes user recommendation to optimization module.The optimization utilizes the both scalar and vector optimization,the scalar optimization utilizes the Bi-objective CF,the vector optimization utilizes the GA-BORF.

3. IMPLEMENTATION

This framework contains the 4 main modules:

- **User Profiles**

As reflected in Fig.1,the MobiContext structure keeps up records of users' profiles for each geographical region.The arrows from clients to venues at lower right of Fig.1indicate the quantity of registration performed by every user at different venues.A user's profile comprises of the user's identification,venues went to by the user,and registration time at a venue.

- **Ranking Module**

On top of users' profiles, the positioning module performs usefulness amid the pre-handling period of information refinement.The pre-preparing can be performed as occasional group occupations running at month to month or week by week premise as arranged by framework administrator.The positioning module applies model-construct HA derivation technique with respect to users profiles to allocate positioning to the arrangement of users and venues in light of shared fortification relationships.The thought is to extracta set of famous venues and master users.We call road as popular,if it is gone by numerous master users,and a user as master if(s)he has gone to numerous prevalent venues.The users and venues that have low scores are pruned from the dataset amid disconnected from the net pre-preparing stage to decrease the online calculation time.

- **Mapping Module**

The mapping module figures comparability diagrams among master users for a given district amid pre-handling stage. The reason for comparability chart calculation is to produce a system of similarly invested individuals who share the comparative inclinations for different venues they visit in a land district. The mapping module additionally processes venue closeness in light of geological separation between the present client and prevalent venues.

- **Recommendation Module**

Fig. 1 delineates the online proposal module that runs a support of get suggestion questions from users. A user's solicitation comprises of: (a) present setting, (for example, GPS area of client, time, and locale), and (b) a limited district encompassing the user from where the top N venues will be chosen for the present client (N is number of venues). The proposal administration passes the user's question to enhancement module that uses scalar and vector streamlining procedures to produce an ideal arrangement of venues. In our proposed structure, the scalar streamlining method uses the CF-based methodology and eager heuristics to produce user favored suggestions. The vector improvement strategy, to be specific GA-BORF.

4. RELATED WORK

Previously, most work concentrated on direction based methodologies for venue suggestion frameworks [1]–[3]. The direction based methodologies record data around a user's visit design (as GPS directions) to different areas, the courses taken, and stay times. The creators in [3] connected information mining and machine learning on direction information to prescribe most prominent spots. In spite of the fact that, direction based methodologies prescribe areas to clients in view of their past directions, a noteworthy downside of such methodologies is that they can't at the same time consider other compelling variables separated from straightforward GPS follow that makes them deliver less ideal proposals. To address such lack, we used multi-target streamlining in our proposed structure. Another issue is that the direction based methodologies experience the ill

effects of information inadequacy issue as for the most part a man does not every now and again visits numerous spots, which brings about meager client venue lattice. Also, the direction based methodologies experience the ill effects of versatility issues as enormous volumes of direction information should be handled bringing on extensive overhead.

A percentage of the methodologies, for example, [3], [5] depend on the online evaluations gave by the clients to the went to puts. The creators in [7] consolidate the accessible venue appraisals with clients' social binds to suggest venues that are high-positioned and in addition most favored by a client's companions. Nonetheless, the creators did not contrast their methodology and any of the standard methodologies, and does not talk about many-sided quality of their work. The previously stated methodologies perform distinctive displaying to clients' inclinations, yet they are not considering different destinations that we particularly considered in our study. In addition, they likewise experience the ill effects of information meager condition issues because of set number of sections inside of the client rating framework.

Aside from rating based methodologies, few of the strategies have their models based on registration based methodologies where the clients give little criticisms as registration about the spots they went to [2]–[4], [7], [14]. For instance, the creators in [6] connected arbitrary stroll with-restart on a client venue registration grid to produce customized suggestions. The majority of the aforementioned approaches have their outlines based on memory-based CF that empowers such ways to deal with give suggestions to clients on the premise of their past sections. Be that as it may, such methodologies experience the ill effects of regular disadvantages of memory-based CF (e.g. cool begin and information sparsity) which decrease their execution. Also, substantial number of likeness calculations on client to-venue lattice makes such methodologies less versatile. There has been some restricted work performed on applying multi-objective optimization on proposal frameworks. One such commitment is by Ribeiro[15] where creators performed a weighted blend of various suggestion calculations and connected improvement to discover proper weights for the constituent calculations. In any case, their methodology is calculation concentrated and no time unpredictability was talked about.

To address the issues refered to above, we proposed a half and half approach over a cloud design that joins the advantages of memory-based and model-based synergistic separating alongside multi-target enhancement to acquire an ideal rundown of venues to be suggested. Besides, our proposed structure shows an answer for adaptability, information meager condition, and frosty begin issues.

5.CONCLUSION

The proposed a cloud-based structure MobiContext cap produces upgraded recommendations by at the same time considering the exchange offs among genuine physical variables, for example, individual's geographical area and area closeness. The noteworthiness and curiosity of the proposed structure is the adjustment of collaborative filtering and bi-objective streamlining methodologies, such as, scalar and vector. In our proposed approach, information inadequacy issue is tended to by incorporating the user to user comparability calculation with certainty measure that evaluates the measure of comparable hobby showed by the two users in the venues ordinarily went to by them two. Besides, an answer for cold start issue is examined by presenting the HA inference model that allots positioning to the users and has a precompiled set of well known unvisited venues that can be prescribed to the new client.

REFERENCES

[1].A.Majid,L.Chen,G.Chen,H.Turab,I.Hussain,and J.Woodward,"A Context-aware Personalized Travel Recommendation System based on Geo-tagged Social Media Data Mining,"International Journal of Geographical Information Science,pp.662-684, 2013.

[2]. M.Ye,P.Yin,and W.Lee,"Location recommendation for location-based social networks," In Proceedings of the 18th SIGSPATIAL International Conferenceon Advances in Geographic Information Systems,ACM,pp.458-461,2010.

[3].Y.Zheng,L.Zhang,X.Xie,andW.Y.Ma,"Mining interesting locations and travel sequences from gps trajectories,"In Proceedings of the 18th international conference on Worldwideweb,ACM,pp. 791-800,2009.

[4]. C.Chow,J.Bao,andM.Mokbel,"Towards Location-Based Social Networking Services,"In Proceedings of the 2nd ACMSIGSPATIAL International Workshop on Location Based Social Networks,ACM,pp.31-38, 2010.

[5].P.G.Campos,F.Díez,I.Cantador,"Time-aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols,"User Modeling and User-Adapted Interaction,vol.24,no.1-2,pp. 67-119, 2014.

[6].A.Noulas,S.Scellato,N.Lathia,and C.Mascolo,"A Random Walk around the City:New Venue Recommendationin Location-Based Social Networks,"In Proceedings of International Conferenceon Social Computing (SocialCom),pp.144-153, 2012.

[7].Y.Doytsher,B.Galon,andY.Kanza,"Storing RoutesinSociospatial Networks and Supporting Social-based Route Recommendation,"In Proceedings of 3rd ACMSIGSPATIAL International Workshopon Location-Based Social Networks,ACM,pp.49-56, 2011.

1. [8].S.Seema,andS.Alex,"Dynamic Bus Arrival Time Prediction,using GPS Data,"In Proceedings of the Nat. Conference Technological Trends(NCTT),pp.193-197,2010.
2.
3. [9].B.Chandra,S.Bhaskar,"Patterned Growth Algorithmusing Hub-Averaging without Pre-assigned

Weights,"In Proceeding of IEEE International Conferenceon Systems, man, and Cybernetics(SMC), pp.3518-3523,2010.

4.

[10].B.Hidasi,andD.Tikk,"InitializingMatrixFactorizationMethodsonImplicitFeedbackDatabase,"JournalofUniversalComputerScience, vol. 19,no.12,pp.1835-1853,2013.

[11].C.ChitraandP.Subbaraj,"Anon dominated Sorting Genetic Algorithm for Shortest Path Routing Problem in Computer Networks,"Expert Systems with Applications, vol.39,no.1,pp.1518-1525, 2012.

[12].Y.Wang,S.Wang,N.Stash,L.Aroyo,andG.Schreiber,"Enhancing Content-Based Recommendation with the Task Model of Classification,"In Proceedings of the Knowledge and Management,pp.431-440,2010.

[13].J.Bobadilla,F.Ortega,A.Hernando,A.Gutiérrez,"Recommender Systems Survey," Knowledge-Based Systems, vol.46, pp.109-132,2013.

[14].J.Bao,Y.Zheng,M.F.Mokbel,"Location-based and Preference Aware Recommendation using Sparse Geo-Social Networking Data,"In Proceeding of 20th International Conference on Advancesin Geographic Information Systems,ACMNewYork, pp.199-208,2012.

[15].M.Ribeiro,A.Lacerda,A.Veloso,and N.Ziviani,"Pareto-Efficient Hybridization for Multi-objective Recommender Systems,"In Proceeding of 6th ACM Conferenceon Recommender Systems,pp. 19-26,2012.

A Novel Identity-Based Encryption using Outsourced Revocation in Cloud Computing

Meenakshi.S[#1], Radhika.B.R[#2], Sripriya.B.S[#3], Tejashwini[#4], Nandini.G[#5]
[1,2,3,4] UG SCHALOR, [5] Assistant Professor, Department Of Computer Science And Engineering,
RAJARAJESWARI COLLEGE OF ENGINEERING, BENGALURU-560074, KARNATAKA, INDIA
meenakshi.yadava94@gmail.com, radhikabr1994@gmail.com, sripriyabs5@gmail.com, tejashwinitejunaik@gmail.com,
nanduamma@gmail.com

Abstract-Public key and certificate management is simplified using Identity-Based Encryption (IBE). Other way to public key encryption is provided by Certificate management at Public Key Infrastructure (PKI). During user revocation the overhead computation occurs at Private Key Generator (PKG) which is the major drawback of IBE. In this project, aiming at tackling the major problem of identity revocation is done for first time by introducing outsourcing computation and propose a revocable IBE scheme in the server-aided setting. During key-issuing and key-updation processes to a Key Update Cloud Service Provider, provides a constant number of simple operations for PKG and users to perform locally. Therefore, this scheme unloads most of the related operations for key generation during issuing and updation of keys. This is achieved by utilizing a novel collusion-resistant technique. Finally, extensive experimental results to demonstrate the efficiency of proposed construction are provided.

Key Terms-Identity based encryption (IBE), revocation, outsourcing, cloud computing.

1 INTRODUCTION

Identity Based Encryption (IBE) is an alternative to public key encryption. Public keys used here are unique name, email address, IP address, etc., Messages are encrypted with client's identity when server uses IBE. Similarly, client receive private key which is equivalent to the identity from Private Key Generator(PKG) to decrypt such cipher text. Any random string can be used as public key in IBE which is a superior advantage over PKI. If some users private key become objectionable then a means to revoke such users from system should be provided. [1]Boneh and Franklin's mechanism would result in an overhead load at PKG (Private Key Generator). Hindering management of certificates is precisely the burden that IBE strives to ease. Boneh and Franklin said that users should update their private keys constantly and senders should use clients' identities combined with present time period. Decriptability can be maintained only if active users update their key in regular manner. Liabilities of named Key Update Cloud Service Provider (KU-CSP) can be overcome in our scheme there is no need of re-issuing the entire private keys instead keys need to be updated constantly.

2 RELATED WORKS

[2]Adel Binbusayyis* and Ning Zhang had proposed Decentralized Attribute Based Encryption Schemes. Existing works toward decentralized ABE can be classified into two categories depending on how the attribute authorities are structured: multi-authority ABE and hierarchical ABE. In the setting of multi-authority ABE, several attribute authorities cooperate to manage the attributes in a system. Each attribute authority is given a unique set of attributes. A user may need to ask more than one authority in order to obtain his/her attributes.

One of the security challenges is how to resist the collusion attack of malicious users. The Chase work achieves collusion resistance by introducing a Global Identifier (GID) given to each user secret key. All the users' secret key components from different authorities will be tied to his GID. However, to make the cipher text be independent of the users GID, a central authority must be used to issue a special secret key for the user using his secret key and the other authorities' secret keys.

[8]Chase and Chow proposed an enhanced multi-authority ABE scheme. They managed to remove the central authority,

but they require that each authority has to assign at least one attribute to each user. This would result a heavy communication cost and a lack of scalability in large scale systems.

[3]Lekwo and Waters proposed a multiauthority ABE scheme that does not require either central authority or cooperation between the multiple authorities. They use a hash function on the user global ID to manage collusion resistance and tie users secret key components together. However, this scheme is not suitable enough to be applied on our scenario because each authority has to know all users GID in advance. In addition, they do not consider how to reduce the workload on an attribute authority when it needs to handle large number of users in large scale system.

[6]Wan et. al. proposed a hierarchical attribute-set-based encryption (HASBE) scheme. Similar to our CP-DABE scheme, their scheme requires a user only to communicate with his/her administering attribute authority, rather than with more than one attribute authority as the case in. However, the HASBE scheme has two drawbacks compared with our CP-DABE scheme. The first one is that our CP-DABE algorithms are faster than the HASBE algorithms in terms of the computational cost. In the decryption algorithm, as an example, in the HASBE scheme, the decryption requires two pairing operations for every leaf node used to satisfy the tree, one pairing for each translating node on the path from the leaf node used to the root, and one exponentiation for each node on the path from the leaf node to the root. However, in our CP-DABE scheme, the decryption algorithm requires only one pairing map for each attribute used to satisfy the access structure. The second drawback is that the HASBE scheme is only proven secure in the generic security model, while our CP-DABE scheme is formally proved against chosen-plaintext attacks under the decisional Bilinear Diffie-Hellman Exponent assumption.

[12]Qingwei Zhang, Mohammed Almulla proposed Revocation Schemes. User revocation is the act of removing privileges from a user so that the user can no longer access the data files. To revoke a user, the data owner needs to update all the attributes keys (i.e. attribute-public keys and attribute-master keys) that have been used to derive the revoked users keys (i.e. user-attribute secret keys). As a consequence of updating the attributes keys, the data owner also needs to update all other affected users keys (i.e. user-attribute secret keys) and re-encrypt each data that have been associated from with any of these revoked attributes.

[13]Pervez, Z. et. al. suggested creating a new access policy with each data file, which contains all the authorized users IDs. To revoke a user, the data owner will only need to remove the revoked users ID from the access policy. This is a straightforward solution, but not practical for large systems since the data owner will need to know the all the users' IDs a priori. To tackle this problem, an expiration time based revocation technique is proposed, which associates an attribute called expiration-time to each user secret key.

3 METHODOLOGY

Implement a portal accessible for two types of users:

- Client
- Server

Clients will be able to look up for the available servers and send a request for adding them to their servers list.

Server will be able to identify the client and either approve/reject the clients ADD request.

Client will use server's public key (IBE) for encryption and sends a message to the server.

Server will be decrypting all clients' message using its private key.

Server will be maintaining a flag called time-to-live for each public key it is going to expose. Once the time-to-live approaches, the KG will auto-revoke all the clients.

Encryption, key generation and user revocation will be handled in a decentralized manner.

4 DESIGN

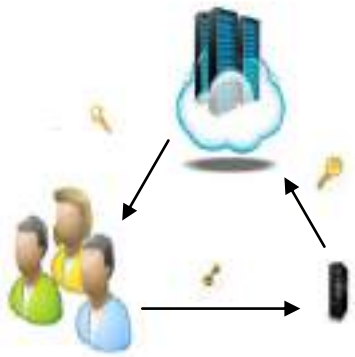


Fig1: Design model for Identity Based Encryption with single KU-CSP

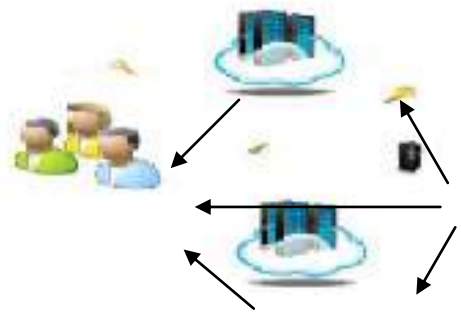


Fig2: Design model for IBE with two KU-CSPs

5 IMPLEMENTATION

Fig1: Front View

Fig2: Login and Register Page

Fig 3: Client Login

Fig 4: Adding Servers

Fig 8: Encrypted and decrypted message

Fig 9: Revocation of Clients

Fig1 demonstrates the front screen of project. Fig2 shows option for client and server to register and Login.Fig3 illustrates client logging into page.Fig4 shows clients2 adding available servers for registration.Fig5 represents number of Registration servers, Pending servers, Expired servers, registered servers names.Fig6 displays the option for server whether to approve or reject client request.Fig7 indicates client sending message to server.Fig8 illustrates the server viewing the message which is both encrypted and decrypted along with sender name and received time.Fig9 demonstrates the server revoking the clients, which can no longer send messages to server.

6 CONCLUSIONS

The approach here is mainly concerned on the issuing of identity based revocation. The mathematical calculation for outsourcing into IBE is introduced and here we are proposing a method of revocable where the operations on revocation are treated as substitute to CSP. By the help of KU-CSP, this method has the following features: a) For both the size of private key at client and calculating the PKG constant efficiency has to be achieved. b) While updating the key the clients are not supposed to contact with the PKG i.e., the PKG is supposed to go offline after the list of revocation to KU-CSP is sent. c) Thorough examination is not needed while updating the key between the client and the KU-CSP. d) Revocable IBE below the stronger adversary model is considerable. Presenting an developed construction and showing that it is protected under RDoC model, where one of the KU-CSPs is expected to be loyal. It is not possible to provide aid to client to obtain his/her de-cryptability again, even if the revoked client and any of the KU-CSPs collude. Lastly, we provide the extensive experimental outcome to demonstrate the efficiency of our proposed construction.

7 REFERENCES

[1] Boneh and Franklin mechanism, R. Schlegel, D. S. Wong, and C. Tang, "A con- ditional proxy broadcast re-encryption scheme supporting timed- release," in Proc. 9th Int. Conf. Inf. Security Practice Experience, 2013, pp. 132–146.

[2] Adel Binbusayyis*, Ning Zhang, and C. Tang, "A CCA-secure identity-based conditional proxy re-encryption without random oracles," in Proc. 15th Int. Conf. Inf. Security Cryptol., 2012, pp. 231–146.

[3] A. B. Lewko and B. Waters, "Decentralizing attribute-based encryption," in EUROCRYPT'11. Springer, 2011, pp. 568–588.

[4] J. Kenney, "Dedicated Short-Range communications (DSRC) standards in the united states," Proceedings of the IEEE, vol. 99, no. 7, pp. 1162– 1182, Jul. 2011.

[5] L. Harn and J. Ren, "Generalized digital certificate for user authen- tication and key establishment for secure communications," Wireless Communications, IEEE Transactions on, vol. 10, no. 7, pp. 2372–2379, Jul. 2011.

[6] Wan et. al, Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization, in Public Key Cryptogra- phyPKC 2011. Springer, 2011, pp. 5370.

[7] M.J. Atallah and K.B. Frikken, ''Securely Outsourcing Linear AlgebraComputations,''inProc.5thACMSymp.Inf.,Comput.Commun. Secur., 2010, pp. 48-59.

[8] C.-K. Chu, J. Weng, Chase, S. S. M. Chow, J. Zhou, and R. H. Deng, "Conditional proxy broadcast re-encryption," in Proc. 14th Aus- tralasian Conf. Inf. Security Privacy, 2009, pp. 327–342.

[9] Q. Tang, "Type-based proxy re-encryption and its construction," in Proc. 9th Int. Conf. Cryptol. India: Progress Cryptol., 2008, pp. 130–144.

[10] L. Ibraimi, Q. Tang, P. Hartel, and W. Jonker, "A type-and-iden- tity-based proxy re-encryption scheme and its application in healthcare," in Proc. 5th VLDB Conf. Secure Data Manage., 2008, pp. 185–198.

[11] L. Cheung and C. Newport, Provably secure ciphertext policy abe, in Proceedings of the 14th ACM conference on

Computer and communi- cations security. ACM, 2007, pp. 456465.

[12] Qingwei Zhang, Mohammed Almulla, Ciphertext-policy attribute- based encryption," in Security and Privacy, 2007. IEEE Symposium

[13] Pervez, Z. et. al Attribute-based encryp- tion for fine-grained access control of encrypted data," in New York, NY, USA: ACM, 2006.

[14] M.J. Atallah and J. Li, ''Secure Outsourcing of Sequence Comparisons,'' Int'l J. Inf. Secur., vol. 4, no. 4, pp. 277-287, Oct. 2005.

[15] A. Sahai and B. Waters. Fuzzy identity based encryption. In EURO- CRYPT, pages 457473, 2005

Portable and Inevitable Path through Versatile Information Gathering in Remote Sensor Systems

Rekha B, SowmyaNaik P.T
Department of Computer Science, City Engineer College, Bangalore, Karnataka
rekhamtechfriends@gmail.com, Sowmya_cec@yahoo.o.in

ABSTRACT-Information gathering is among issues continually obtaining consideration in zone of remote sensor systems (WSNs). There is a reliable increment in examination coordinated on additions of applying versatile components (MEs) to gather information from sensors, particularly those situated to power issues. Two winning systems used to gather information in sensor systems. Main methodology requires information bundles to be overhauled by means of multi-jump hand-off to achieve the particular base station (BS). Sensor sends their packets through other middle sensors. Information created from the sources in field is frequently repetitive and very associated. In like manner, assembling and collecting information from the area in the sensor systems is essential and important to spare the vitality and remote assets of sensor hubs. Present the idea of a neighborhood sink to address this issue in geographic steering. Neighborhood sink is a sensor hub in area, in which sensor hub is briefly chosen by a worldwide sink for get-together and amassing information from sources in district and conveying totaled information to worldwide sink. Second approach includes a ME which serves as the center component for the seeking of information. Mobile Element will visit transmission scope of every sensor to transfer its information before in long run coming back to BS to finish the information transmission. This methodology has demonstrated to diminish vitality utilization considerably when contrasted with multi-bounce technique. Mobile element will choose a centroid point between two sub-surveying focuses, consequently selecting regular defining moments as center of the premise of visit way. Broad discrete-occasion reproductions have been created to evaluate execution of proposed calculation. Proposed calculation keeps up the power utilization inside of an adequate level.

Keywords—Wireless sensor networks, Aggregating, Global sink, and Local sink, Mobile data gathering; Polling points; Turning points; Mobile Element, Base Station, mobile sinks

1. INTRODUCTION

Most existing geographic steering conventions on sensor systems focuses on discovering approaches to ensure information sending from the source to the destination, relatively few conventions have been done on social event and amassing information of sources in a neighborhood and contiguous district. Numerous geographic directing conventions have been proposed to course effectively information in remote sensor systems on the grounds that these remote sensor systems are conveyed by various sensor hubs and sensor hubs are alterable. In any case, most existing geographic steering conventions [4, 9] on sensor systems focuses on discovering approaches to ensure information sending from the source to destination, very few conventions have been done on social occasion and conglomerating information of sources in a nearby and contiguous area where creates information concentrated by clients. In any case, information gathering and total of sector in remote sensor systems is critical and important to spare vitality of sensor hubs [7, 5, 8], in light of fact that information created from sources in district might be regularly repetitive and exceptionally associated [11, 10] since remote sensor systems are thickly sent by an expansive number of sensor hubs [7]. It however produces different issues that a worldwide sink gathers and totals information created from domain as appeared in Figure 1(a). Firstly, vitality utilization increments with quantity of source hubs in light of fact that their information is spread to worldwide sink without the accumulation. Furthermore, same number of information are scattered by means of comparable ways in geographic steering strategies, sensor hubs on the ways devour much vitality and experience the ill effects of information blockage. Thirdly, if there are information in different areas keeping in mind end goal to be dispersed on comparative ways, they are hard to be scattered attributable to the information clog [9]. Consequently, it can decrease vitality utilization and information blockage that information created in zone are gathered locally in district and accumulated information are dispersed to global sink, as indicated Figure1(b).

To minimize the repetition of information from sources in physical vicinity and henceforth diminish the information movement in remote sensor systems, viable informationassembling and conglomerating instruments [3, 5, 6] have been proposed. Information assembling and amassing approaches [3, 5] in level systems have been proposed in tree-based structures, in which information total happens sharply when their courses converge. To give versatility and vitality productivity, various leveled information assembling and collecting approaches [7, 8] have been proposed in bunch based structures, in which information from various sources in a group is totaled at an uncommon hub called a bunch head. In like manner, since these two methodologies consider the entire sensor fields rather than the areas of sources, they spend numerous vitality and remote assets to develop and keep up the trees and the bunches in locales which don't produce sources.

2. RELATED WORKS

In this segment, we depict a system model to execute our work. Sensor hubs are conveyed in the sensor system. A Global Sink situates in a particular position of the system. A

worldwide sink is a sink which gathers information from the entire sensor fields and conveys them to clients in engineering of general remote sensor systems [1]. Every hub knows about its own area through accepting GPS signals or through restriction methods [2]. All sensor hubs can know region of worldwide sink by programming the area to sensor hubs or flooding the area by the global sink. Every sensor hub can know about its own particular remaining vitality. We characterize a Target Region as the particular locale where is produced information from numerous source hubs by creating of occasions or is intrigued by a client. On the off chance that sensor hubs in an objective district distinguish an occasion, they get to be source hubs and produce information of same size for the occasion. At that point, the source hubs incorporate their won field data in their information since numerous applications in remote sensor systems require area of source information, for instance, target following and natural surroundings observing, furthermore incorporate their won remaining vitality which is measured by force generator. Next, source hubs disperse their information to the worldwide sink by geographic directing [4, 9]. In wake of accepting information with area data from source hubs in an objective district,worldwide sink can choose one among general sensor hubs in objective district to work as a Local Sink. The nearby sink gathers and totals locally information in objective district and conveys collected information to a worldwide sink. We introduce techniques to decide single neighborhood sink and different nearby sinks to bolster versatility as far as the cushion size of neighborhood sink and due date of information.

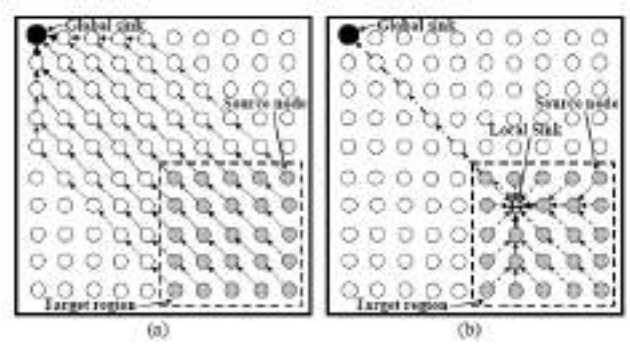


Fig 1: Data gathering of two methods in a geographic routing(a)datagathering without a local sink and (b) data gathering with a local sink

Keeping in mind the target goal to partition an objective zone by 4 various nearby sinks, Figure 2 demonstrates an illustration of a circle that is straightforward as far as time many-sided quality. 360 level of the circle is partitioned by 4 and every neighborhood sink take charges a sub target region of 90 degree in circle target locale. A nearby sink is distributed in every sub target district, and accumulate information from its sub target area and disperse collected information to a worldwide sink. To start with, power utilization increments drastically when multi-bounce information gathering methodology is connected. Second, dormancy increments when the versatile information gathering methodology is connected through a solitary bounce.

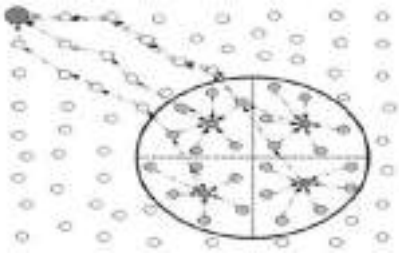


Fig 2: Circle to divide a target region into 4 equal size parts

Be that as it may, noteworthy vitality is spared when a suitable information gathering methodology is connected. Consequently, expanding the quantity of hubs that are gone by topic causes a long visit way which infers expanding dormancy. It is, subsequently, evident that there is an inborn exchange off between vitality utilization and inertness in relationship to the properties of the ME.

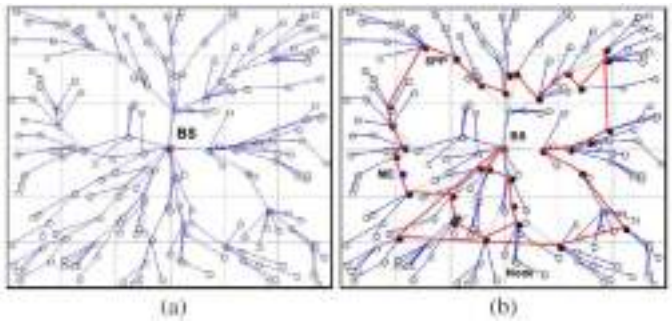


Fig 3 : (a)Multi-hop data gathering. **(b)** Mobile data gathering.

Figure 1 represents the procedure of pre-determined information gathering approaches. A system with 200 sensors is conveyed arbitrarily with static BS situated at focal point of a 200 m × 200 m organization field. Figure 3(a)multi-jump information gathering approach by embracing multi-bounce directing. In this approach, every parcel is sent through different sensors to achieve BS utilizing system of most brief way with a base jump number. The portable information gathering approach in view of sub-surveying focuses (SPPs) is delineated in Figure 3(b). Every sensor sends the parcels to closest SPP and anticipates movable element to transfer them when going to the individual SPPs. The visit way of ME that visits each SPP and static BS is spoken to by a red strong line[13]. This is because of need to just visit chose focuses, in this way minimizing the vitality utilization of every sensor because of limited neighborhood information accumulation to the SPPs. Mobile agent known as Sensor Car (SenCar), is furnished with two reception apparatuses that can transfer information simultaneously from two perfect sensors by means of short-range correspondence. In any case, even with lessened information transferring time, this plan unfavorably draws out visit length, particularly when utilizing one and only SenCar. This is ascribed to prerequisite that SenCar ought to be inside of the transmission scope of every sensor hub. SenCar must visit some particular areas where more sensors

are perfect, which may prompt expanded visit length. In this way, the ideal arrangement was recommended to be a harmony in middle of SDMA and a shorted visit length of SenCar. Rather than unconstrained information gathering time in [12], creators in [1] proposed information accumulation utilizing different versatile hubs. These portable hubs gather the information from every sensor each t second to maintain a strategic distance from information flood because of restricted storage of the sensors

3. NETWORK MODEL

3.1 Shortest Path Tree-based DataGathering(SPT-DGA)

SPT-DGA Shortest Path Tree-based Data Gathering procedure, primary work is to fuse limited multi-hop transfer into versatile information gathering. The procedure to choose certain hubs as surveying point hubs is by building a most limited way tree to closest hub to BS. Calculation chooses surveying point hubs taking into account the most remote leaf hub and the jump check is given. The induction of surveying focuses taking into account SPT-DGA is as takes after every sensor must partner to one and only surveying point association with the upper layer to guarantee tree structure aside from surveying point which ought not have any association with upper layer and each surveying point has at most d levels.BS is considered as an uncommon surveying point which assembles information from portable authority as it were. The SPT-DGA procedure has numerous restrictions with respect to minimizing visit length of ME in view of surveying based methodology. These constraints include fabricating tree to closest hub to BS and surveying focuses being covered. Subsequently, visit way increments on grounds that ME visits each surveying point independently in arrangement field. Figure 4 (a) presents 50 hubs dispersed arbitrarily over arrangement field whilst BS is situated at focal point of field. Usage of SPT-DGA calculation will bring about ten surveying focuses including BS as delineated in Figure 2b. Figure 4 (b) demonstrates that hubs (9, 10), (13, 26), (31, 35) and (15, 49) which have been chosen as surveying focuses are covered. This is because of conduct of SPT-DGA calculation to choose surveying guides which to be sure lead toward cover between surveying focuses.

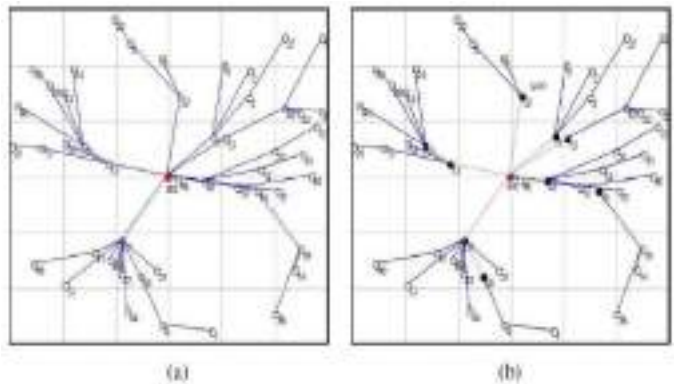


Fig 4: (a) Distribution of sensor nodes. (b) Distribution of polling points.

3.2 Mobile Data Gathering-based Network Layout

Two issues are genuinely considered in mobile information gathering. They are the force utilization of every sensor and visit length, which influences idleness brought about because of information social occasion. Crossing all sensor hubs by utilizing ME is not a favored arrangement since information gathering inactivity increments because of restricted speed of ME. Then again, information which has crossed various jumps before achieving ME/BS faces higher vitality utilization. Striking a harmony between these two issues is essential thought in this paper. In accomplishing this, procedure is isolated into two stages. Primary stage is to discover a subset of sensor hubs called sub-polling point(SPPs).These focuses assume part of neighborhood information collection from every single associated sensor. Nearby information collection is controlled by quantity of bounces including crossed information which depends on application needs. Second stage is to locate a typical point which is covering between two SPPs, which ought to be known as CTP. These covering CTP are chosen in light of particular system format (i.e. restriction).

CTPs are considered as delay areas for the ME to transfer information bundles which are supported at SPPs through single bounce. ME begins information gathering visit way from BS and navigates through all CTPs, gathering information bundles from particular SPPs and after that in long run coming back to BS. Figure 5 shows the portable information gathering, where sub-set of sensor hubs chose as SPPs, meant by dark filled circles, totalinformation from associated sensors. ME visit way depends on the SPP areas as appeared in Figure 5 (a). Determination ofway should be pondered in following segment. Figure 5 (b) shows the CTPs chose in view of SPPs spoke to by red circles. Reaching all CTPs and BS is viewed as last visit way of ME.

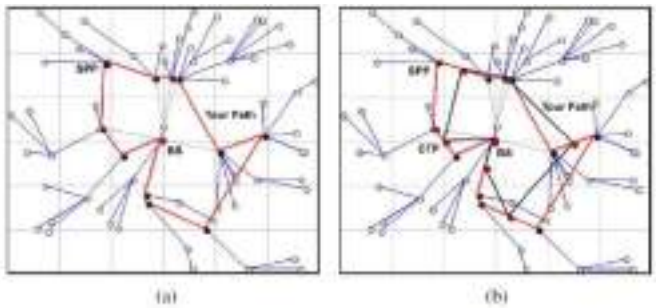


Fig 5: (a) Mobile data gathering based on SPPs. (b) Mobile data gathering based on CTPs.

4. Architecture of ME tour path based on SPPs and CTPs

Main stage is to discover a subset of sensor hubs called sub-polling point (SPPs).These focuses assume part of nearby information accumulation from every partnered sensor. The nearby information conglomeration is controlled by quantity of jumps including crossed information which depends on application needs. Second stage is to locate a typical point which is the covering between two SPPs, which ought to be known as CTP. These covering focuses CTPs are chosen in view of separate system design (i.e. confinement). CTPs are

considered as interruption areas for ME to transfer information parcels which are supported at SPPs by means of single bounce. ME begins the information gathering visit way from the BS and navigates through all CTPs, gathering information parcels from separate SPPs and afterward in long run coming back to BS.

The proposed procedure will profit by aftereffects of to minimize the visit way of ME. Finding ideal area of sub-set of sensors known as SPPs, directing with the briefest way and visit length of versatile information social affair ought to be tended to in a brought together way to upgrade portable information gathering dormancy. As examined before, so as to locate ideal area of CTPs among SPPs, most limited way directing and visit way of ME ought to be together considered. From one viewpoint, when no CTP is accessible, for each SPP, most ideal approach to gather information is by going by each SPP independently by ME, under the supposition that inactivity of information social event is corresponding to speed of ME. Then again, when CTP is accessible, visit length is adequately abbreviated in two ways. To begin with, BS considers a CTP which accumulates information from sensors found adjacent it. Second, the visit way experiences the CTPs which are littler than quantity of SPPs. Mobile Data Gathering-based Network Layout (MDG-NL) with its pseudo code is recorded. The fundamental thought of this calculation is to discover a covered point between two SPPs in which latency of information collecting improved. The comprehension of calculation is further expounded in Figure 6 which delineates procedure of sending sensor hubs until social occasion of detected information from all sensor hubs. Thirty sensor hubs are scattered on 25 m × 25 m arrangement field with static BS put amidst field. Furthermore, quantity of transfer jump number has been limited to two bounces most extreme. Figure 6(a) speaks to 30 hubs consistently irregular dispersed and built as a most brief way tree over the arrangement field to the BS as a root. Four SPPs are determined including static BS (i.e. 7, 12, 21 and BS).

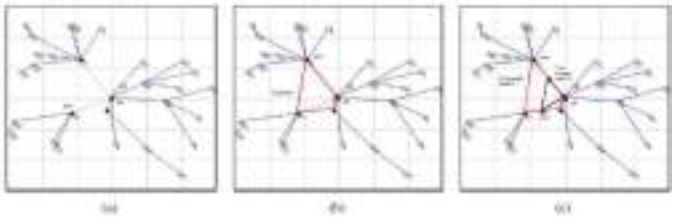


Fig 6 :ME tour path based on SPPs and CTPs. (a) Constructed SPT and derived SPPs. (b) Mobile data gathering based on SPPs. (c) Mobiledata gathering based on CTPs.

Figure 6 (b) delineates information gathering visit which begins from static BS, goes through all SPPs recorded above and after that inevitably comes back to BS. In illustration above, computing the Euclidean separation that interfaces all SPPs including the BS as beginning and closure focuses as delineated in Figure 6 (b) (i.e. BS → 7 →12→ 21→ BS) results in 13.30 m. Moreover, as specified prior in Section 1, the pace of ME is around 0.1 to 2.0 m/s. Considering that 1 m/s is normal rate of ME, the time expected to complete information gathering visit way is gotten by isolating aggregate separation over the normal speed (i.e. 13.30m/1s). Subsequently, MEneeds around 13.30 s to achieve all SPPs barring information transferring time. Figure 6 (c) shows the CTP approach which is spoken to by three CTPs just, including the BS. Furthermore, regions of chose CTPs are nearer to BS

5. PERFORMANCE EVALUATIONS

Execution of SPT-DGA and MDG-NL as an element of limited hand-off bounce d as far as aggregate vitality devoured n Figure 7(a). It is clear that when d has the littlest worth, aggregate vitality expended is minimized in both calculations because of minimized weight at every sensor hub to convey other information. Clearly, when d expands, the aggregate vitality devoured increments as well. This is because of different sending information bundles through hubs. Aggregate power expended utilizing MDG-NL is not exactly vitality devoured utilizing SPT-DGA. Power utilization at a specific level while minimizing the visit length of ME, is a test because of the exchange off between vitality utilization and visit length in versatile information gathering [1].Figure 7(b) shows execution of SPT-DGA and MDG-NL as a component of transmission reach Tr as far as aggregate vitality expended. It is evident that when transmission range Tr has littlest quality, aggregate vitality expended is minimized, and this is because of two reasons.

To begin with, force utilization because of correspondence is influenced specifically separation.Second, various detached systems are made which prompts expanding quantity of SPPs and CTPs with a couple of sensors partnered. As it were, level of each geometric tree is constrained and some of time there is one and only level. Moreover, expanding transmission range constrains sensors to send their information to most remote neighbor towards BS. In this way, chain of command level of most brief tree is diminished with making a couple surveying hubs (i.e. SPPs and CTPs). In both calculations, aggregate vitality expended is verging on like each other. Figure 7 (c) outlines execution of SPT-DGA and MDG-NL as a component of quantity of hubs N regarding all out vitality expended. It is seen that when N has the littlest quality, aggregate of sensor hubs subsidiary to each SPP is less. Along these lines, correspondences required to send information to closest SPP among leaf hubs, when sensor hub N builds, aggregate vitality devoured increments too because of expanding number of created bundles which prompts expanding the quantity of interchanges required to convey information to closest SPP. In both calculations, aggregate vitality expended is nearer to each other. Figure 7 (d) represents execution of SPT-DGA and MDG-NL as an element of sent field L as far as aggregate vitality devoured. It is seen that when conveyed zone L has littlest quality (i.e. 100), force utilization is generally most astounding in contrast with different qualities. This is on account of sensor system is completely associated and all correspondences required to send the information are registered. Then again, when L has the most astounding quality (i.e. 500), force utilization isminimized. This is on grounds that various detached systems are made with a couple of sensors. Accordingly, correspondences required to convey information to closest

SPP areminimized. Likewise, a few sensors are situated far fromany other system. Notwithstanding, MDG-NL keeps up force utilization inside of a specific level, and force utilizations for both calculations are comparable.

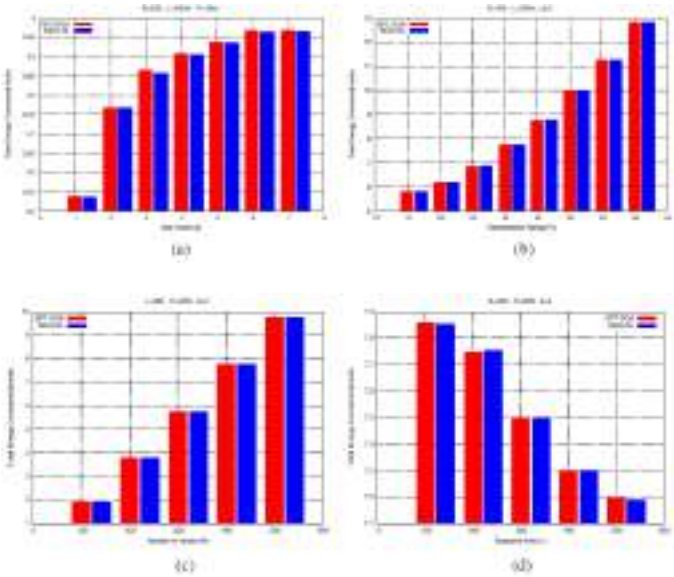


Fig7 : SPT-DGA vs. MDG-NL in terms of total energy consumption. (a) Hop count. (b) Transmission range. (c) Number of nodes. (d) Deployed area size.

6. CONCLUSIONS AND FUTURE WORK

In this exploration, an itemized depiction of versatile information gathering inWSNs taking into account defining moments has been talked about. Two created procedures to improveinformation get-together are exhibited. In principal calculation, detected information are accumulated utilizing a multi-jump way to deal with specific hubs called SPPs which are limited by a specific level. The bouncing is for obliging the force utilization by restricting the interchanges among hubs. Second calculation, ME chooses quantity of CTPs to pull the information from two SPPs at one respite. This helps ME to minimize visit length and idleness to convey information to BS. Created approach has demonstrated that it defeats SPT-DGA approach which empowers social occasion of information from one and only surveying point at every interruption.

This is because of disposal of superfluous visit way required to visit each surveying point independently. Broad reenactment was performed to approve created calculations and to consider effect on execution measures in contrast with SPT-DGA calculation. MDG-NL has demonstrated to effectively adjust and essentially enhance visit length of ME and inactivity of information social event. In any case, because of tradeoff between force utilization and visit length of ME, MDG-NL keeps up force utilization to be inside of a worthy level in contrast with SPT-DGA calculation. Upgrade of applying numerous MEs with locale division is a fascinating range later on. With this improvement, every ME is named to a predefined sub-district, which is a part of arrangement field.

REFERENCES

[1] C Wang, H Ma, Data collection in wireless sensor networks by utilizing multiple mobile nodes, in Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN '11) (Beijing, 16–18 Dec 2011), pp. 83–90

[2] N. Bulusu, J. Heidemann, and D. Estrin, "Gps-less low cost outdoor localization for very small devices," IEEE Pers. Commun., Oct. 2000.

[3] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," inProc. ACM MOBICOM, Aug. 2000.

[4] B. Karp and H. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," in Proc. ACM MOBICOM, Aug. 2000.

[5] C.Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, "Impactof Network Density on Data Aggregation in Wireless Sensor Networks," in Proc. IEEE ICDCS, Jul. 2002.

[6] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," IEEE Trans. Wireless Commun., Vol. 1, No. 4, pp. 660–70, Oct. 2002.

[7] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," IEEE Communication Magazine, Aug. 2002.

[8] O. Younis and S. Fahmy, "HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor networks," IEEE Trans.Mobile Comput., vol. 3, no. 4, pp. 366–79, Dec. 2004.

[9] T. He, J. Stankovic, C. Lu, and T. Abdelzaher, "A Spatiotemporal Communication Protocol for Wireless Sensor Networks," IEEE Trans. Parallel Distrib. Syst., Vol. 16, No. 10, pp. 995-1006, Oct. 2005.

[10] R. Rajagopalan and P.Varshney, "Data-Aggregation Techniques in Sensor Networks: A Survey," IEEE Commun. Surveys Tuts, Vol. 8, No. 4, pp. 48-63, 4th Qua. 2006.

[11] K. Yuen, B. Liang, and B. Li, "A Distributed Framework for Correlated Data Gathering in Sensor Networks," IEEE Trans. Veh. Technol., Vol. 57, No. 1, pp. 578-593, Jan. 2008

[12] M Zhao, M Ma, Y Yang, Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks. IEEE Trans. Comput. **60**(3), 400–417 (2011)

[13] M Zhao, Y Yang, Bounded relay hop mobile data gathering in wireless sensor networks. IEEE Trans. Comput. **61**(2), 265–277 (2012)

ADVANCED NETWORK MONITORING SYSTEM USING A HYBRID OF AGENT AND AGENTLESS CONCEPTS

Rahul S Sreedhar, Prithvi Vihari R, Dr. N Guruprasad
[1,2]UG student, [3]Professor,[1,2]CMR Institute of Technology,[3]New Horizon College of Engineering, Bangalore
rahuls.1234sreedhar@gmail.com , prithvivihari11@gmail.com, nguruprasad18@gmail.com

Abstract-Network monitoring system on agent based protocol provides benefits such as better security, increased bandwidth efficiency and increased monitoring capabilities. However in today's world, an agentless network monitoring system would seem to be more apt as it would reduce the cost of installation & maintenance, reduced administrative overhead, fast detection of network outages and protocol failures and is easy to deploy. For large organizations, factors such as cost of installation, reduced deployment time and maintenance are as critical as security. Hence, this paper would compare the concepts of both agent and agentless monitoring protocols and discusses the best ways to combine and reap their benefits.

Keywords— Agent-Based, Agent less, NagiOS, RT Ticket Management System, Simple Network Management Protocol

I INTRODUCTION

Data communications and networking influences the way business are done and the way we live today. Quicker business decisions are required to be made which requires immediate access to accurate information for decisions makers. Business today relies more often on computer networks and inter-networks. Larger organizations reply more on big network topology to avoid the manual networking options that causes waste of resources and that could lead to efficiency improvement and has become best industry practices to date.

Conventional Network management systems involves monitoring, testing, configuring and troubleshooting network components to meet smooth and efficient operation of network that provides adequate quality of service for the users or meet the organizational requirements. This task is accomplished generally through a most common management system, the Simple Network Management Protocol (SNMP) concept that leverages on the use of hardware, software and humans. SNMP uses the concept of manager and agent, with the former being usually a host, who controls and monitors a set of agents, usually routers. The application level protocol is so designed so that it can monitor devices made by different manufacturers and installed at different physical networks [7].

Over the years, various updates of SNMP have been released. As of date, the SNMPv3, have progressively increased security features such as Authentication, Remote configuration and administration capabilities Privacy, Authorization and access control, that were lacking in the previous versions [7]. Nevertheless, the notable deficiency in the difficulty in monitoring networks as opposed to nodes on networks, improved ease of operational efficiency, performance and reduced infrastructure cost for deployment continues to be challenging while adopting the agent based concept.

The concept of agent and agent-less network monitoring, as shown schematically in Figure 1, of various network devices without the need to install software agents across all the monitoring devices, appears to be an attractive option to overcome few of these challenges. However, it is noted that the agent less options does have major deficiencies in respect of security aspects.

Figure 1: Factors affecting network monitoring

This paper explores a hybrid approach which proposes to exploit advantages of both the Network Monitoring Concepts towards four focal areas such as ease of deployment, adaptability and less node on network and security aspects.

a) AGENT BASED SYSTEM

Monitoring with agents has the cost of installation, configuration (proportionate to number of managed elements), platform support needs and dependencies. It involves installing the agent software in all the systems connected to the network.

In general, agent-based monitoring is regarded more secure since it does not necessitate storage of passwords centrally. Organizations that have high-level of security compliance needs would prefer agent-based monitoring as a result.

In each node system of each sub-network, the Agent Processor, which processes management information based

on a management information model, processes the information of each of the equipment by each management service and stores the situation information of recent equipment and network as an object-oriented model in the Management Information Base. The center operating system analyzes the relations among the sub-network units.

There are crucial systems that support SLA's (Service Level Agreements) for 99.999% (5-nines) availability. It is important to monitor during each minute and second, how the system is performing – the peaks and trough of usage, not just average utilization levels. This is crucial for capacity analysis and prediction heuristics. While it is theoretically possible to poll at shorter intervals than a minute, it is not efficient to go to really low intervals using remote polling approach (as minor delays due to network latency can cause havoc with the data collections).

This system enables in-depth monitoring and management. The agents to application / OS (Operating System) communications are handled internal to the server. Hence, no additional firewall rules need to be configured since it is much more secure.

Figure 2: Agent Based Monitoring Architecture

Agents need to be deployed on each server. The data is collected locally and only the processed final results are transported to the console and therefore the bandwidth is utilized efficiently. Any software vendor will tell you that their agents work the best with their platform. While this may be true, it may also be because their management platform is built only to work with their proprietary agents. The result is vendor lock-in, and changing vendors can mean expensive, large-scale, long-term deployments of replacement technology. Because of this, when IT requirements change, it can be extremely costly to meet them. Generally, open standards and flexibility work far better in the long run.

b) AGENTLESS SYSTEM

Agentless monitoring is easy to deploy, as the monitoring and configuration happens at a central place with a good UI. Costs of installation (and the ongoing maintenance) are really low.

Considering one of the case studies of HP global financial service incorporation, when the company began reconsidering its monitoring technology, it identified a number of objectives. One was cost. "We wanted to reduce the amount of money we were spending on agent maintenance and support", notes William Gillen, Director of Systems Engineering, global financial services corporation. "At the time, I was helping to support an environment with about 12,000 servers, and we recognized that managing our monitoring agents was increasing our support costs by 40%." [6].

In reality, all computer tasks require related programs to run and these programs could be considered agents. However, operations can generally be performed on a controlling machine that may use an agent, while its target is agentless in that it need not install or run new software related to the task itself. This capacity can save time required for managing agents on numerous target machines, especially in large enterprises. That said, even some software called agentless may use built-in services in an operating system, and they may require configuring. So the term agentless can be a bit misleading.

Agentless software generally requires the creation of a user account on the target machine or machines to facilitate access to the data on the account related to the desired operation. Agentless programs directly access the files, often remotely, via this user account. So the created profile must have the necessary access to these files and the software must store the login credentials to access the profile. This access may be facilitated through a number of different Internet communication standards (e.g., FTP (File Transfer Protocol), TELNET, SSH (Secure Shell)).

For agentless monitoring, implementation ranges from built-in SNMP agents to remote shell access, such as SSH. "Agentless" is a bit of a misnomer. All management requires an agent, whether the agent is embedded in the management platform, the managed device, or a separately installed piece of software. The industry has accepted the definition of agentless as a management agent that is embedded in the software of the device or as a capability of the manager, requiring no separate installation or licensing.

Agentless monitoring really means the use of existing, embedded capabilities.

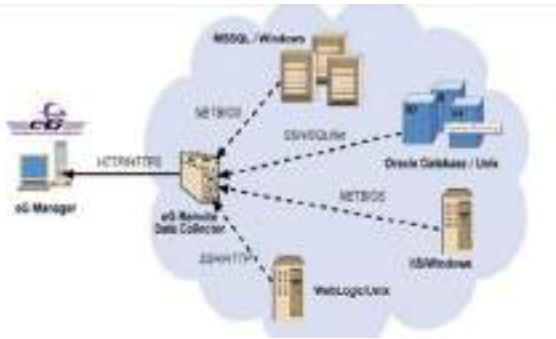


Figure 3: Agentless Network monitoring Architecture

II. NETWORK MONITORING BY NAGIOS

Nagios is open source and web based software used for Network monitoring [1]. It monitors network nodes and services applied on them and inform the network administrator when any change happens in the network [2]. Nagios is well suited application for Linux environment but it can also run on other platforms as well. Nagios is a secure

and easy manageable application which provides a good web interface, automatic alerts if condition changes and various notification options [3]. When any node or service in the network faces a problem, Nagios generates notification to the network administrator in the form of email or SMS. Nagios is developed under GNU general public license and supports different services like HTTP (Hyper Text Transfer Protocol), NNTP (Network News Transfer Protocol), Ping, SMTP (Simple Mail Transfer Protocol), etc. Nagios allow administrator to build complete network topology and define child-parent relationship among nodes. This child-parent relationship among nodes enable Nagios to send only one notification if a parent node goes down with the information that child nodes become unavailable. A generic network topology created in Nagios is shown in Figure 4.

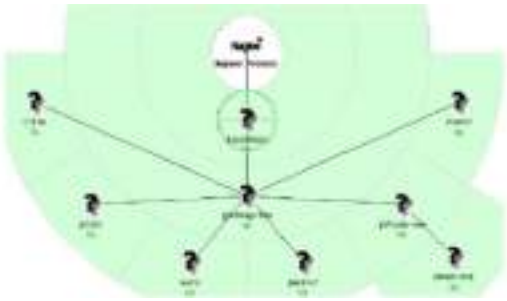


Figure 4: Generic topology created in Nagios

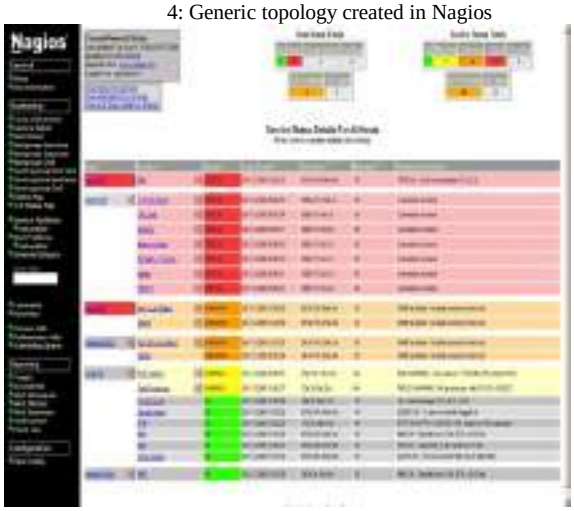


Figure 5: Service status details for all hosts [8]

Nagios decide about the condition of nodes and services with two factors: status and type of state. The status can be either up, down, critical or unreachable while the type of state can be either soft state or hard state. The type of state has great importance for alerting process. It decides about the final status before a notification is sent out. In order to avoid false notifications, Nagios check the nodes and services for pre-defined number of times before declaring them to have real problem [3]. The number of attempts can be controlled by max_check_attempts option in the node and service definitions. Node or service is declared in soft state if status check results in a non-OK state but the number of attempts is less then max_check_attempts. This is also called „soft error state. In the soft recovery state, the node or service recovers from „soft error state". Node or service is declared in hard state if status check results in a non-OK state for the number of attempts specified in max_check_attempts. This is also called hard error state when the node is either unreachable or down. In the hard recovery state, the node or service recovers from ard error state. The hard state of node or service will change if the status check changes from hard OK state to hard non OK state or vice versa. If during the hard state change, the node or service is declared in non-OK state then the hard node or service problem is logged and administrator is notified about the problem. But if during the hard state change, the node or service is declared in OK state then the hard node or service recovery is logged and administrator is notified about the recovery. Furthermore, if the hard state change occurs from one non-OK state to another non-OK state then the administrator is re-notified about the problem [1].

a) NAGIOS INSTALLATION

Nagios installation in Ubuntu is quite simple. Just follow the steps below as root user: # apt-get install nagios3. After installation, assign the web user password with the following command:

```
# htpasswd -c /etc/nagios3/htpasswd.users username
```

Designed with scalability and flexibility in mind, Nagios gives you the peace of mind that comes from knowing your organization's business processes won't be affected by unknown outages.

Nagios is a powerful tool that provides you with instant awareness of your organization's mission critical IT infrastructure. It allows you to detect and repair problems and mitigate future issues before they affect end-users and customers. It is a popular open source computer system and network monitoring software application. It watches hosts and services, alerting users when things go wrong and again when they get better. It also monitors your entire IT infrastructure to ensure systems, applications, services, and business processes are functioning properly.

In the event of a failure, Nagios can alert technical staff of the problem, allowing them to begin remediation processes before outages affect business processes, end-users, or customers. With Nagios you'll never be left having to explain why a unseen infrastructure outage hurt your organization's bottom line.)

Now Nagios can be accessed from browser using Fully Qualified Domain Name (FQDN) by visiting the web page at http://FQDN/nagios3.

Use the login information specified above:
username: username
password: password

b) NAGIOS CONFIGURATION

Make the network topology by defining every node in /etc/nagios3/conf.d/ directory. File name should be the same as host_name. A generic node1 can be defined as follows:

```
define host {
use        generic-host
host_name        node1
alias      node1 in network
```

```
address     [node1 IP address]
parents     node1's parent if any
}
```

Introduce a group in hostgroups_nagios2.cfg that will include all above defined nodes

```
define  hostgroup {
hostgroup_name   network-group
alias      network nodes
members       node1, node2,…
}
```

Associate some services e.g, ssh, ping etc to the defined group in „services_nagios2.cfg file.

```
define  service {
hostgroup_name     network-group, othergroups
service_description  PING
check_commands check_ping!100.0,20%!500.0,60%
use    generic-service
notification_interval     0;
for re-notification,
set > 0
}
```

Since nagios has to be interfaced with RT for better network management, therefore define RT contact in the file contacts_nagios.cfg.

```
 define contact {
contact_name   RT
alias  RequestTracker
service_notification_period  24×7
host_notification_period  24×7
service_notification_options    c,w,r,u
host_notification_options r, d
service_notification_commands   notify-serviceby-email
host_notification_commands    notify-host-byemail
 email       rt@host.FQDN
}
```

Now introduce a contact group that will include all defined contacts

```
define  contactgroup {
contactgroup_name      Network-admins
alias      Network and Nagios admins
members       RT, other-contacts
}
```

Finally, restart the nagios and check for the applied configurations in the web interface.

```
 #  /etc/init.d/nagios3  restart
```

III. RT TICKET MANAGEMENT SYSTEM

A good management system is usually required for organizations in order to manage their work flow, offering services to clients or manage hardware/software problems. Every ticket has certain attributes and ID number used to identify the ticket. RT is open source ticket management software developed by Best Practical, Inc. New-York University. RT is heavily used worldwide as it provides email friendly interface and keep track of tickets which represent a job to be done. RT provides ease of use, multiuser accessibility, access control, history tracking and remote accessibility; generate notifications and customization according to organization requirements.

Different versions of RT software are available to work on windows, UNIX and Linux environments. It also requires a database which can be MySQL, POSTGRESQL or ORACLE. Since RT is open source, thus can be customized using Perl script language. RT also requires Apache web server. RT makes use of Perl based main engine and a database to store its data and provides web and email interfaces.

RT allows creating different users via web interface and assigning rights to them. Users can also be arranged in groups and assign rights to them on global basis. RT can also be configured to generate queues of tickets to work on. These queues correspond to a group of different services.

Ticket is key object in RT which defines a job to be done. RT ticket attributes include status, watcher, time left, time worked, ticket priority, queue and its owner. Main ticket watchers are its owner and requester but additional watchers can also be defined. RT ticket priority can range from 0-99 which determines the importance of ticket with 99 as the highest priority. It is also possible to define initial and final ticket priority which increases or decreases with the time left. RT also allows defining custom scripts which take an automatic action in response to a given condition.

a) RT INSTALLATION

RT software requires many dependencies for its installation in Linux environment. Ubuntu and Fedora are preferred choices as they allow automatic installation of many required dependencies during the installation process. RT can be installed in Ubuntu environment with the following commands in the terminal:

```
 # apt-get install rt3.6-apache2 request-tracker3.6
3.6clients apache2-doc postfix mysql-server lynx libdbd-
pgperllibapache-dbi-perl  rt3.6-rtfm
```

During postfix configuration, a pop up window appears to enter the „system mail name‟ which is also called Fully Qualified Domain Name (FQDN). FQDN is used to provide global access to the RT software.

b) RT CONFIGURATION

Make the following important changes in the configuration file „RT_SiteConfig.pm. Set($rtname, „rt-name‟);

```
Set($Organization,„organization-name‟);
Set($CorrespondAddress , rt@FQDN);
Set($CommentAddress , rt-comment@FQDN);
Set($WebPath , "/rt");
Set($WebBaseURL , "http://FQDN/rt");
Set($DatabaseType, $typemapmysql); Set($DatabaseUser ,
„user-name‟); Set($DatabasePassword , „user-password‟ );
```

Now restart the apache to get sure that all the changes have been recorded. Enter the following URL in the browser: "http://FQDN/rt" and finally, log in with the user name "root" and password "password".

IV. INTERFACING NAGIOS WITH RT

At the final stage, Nagios is interfaced with RT software. The main network monitoring task is performed by Nagios but the ticket management task is performed by RT. „rtmailgate" plays an important role for creating interface. For this purpose, an alias is created in file called, aliases by inserting the following text:

rt: "|rt-mailgate --queue `name of RT queue' --action correspond --url http://FQDN/rt"
Rt-comment: "|rt-mailgate --queue `name of RT queue' -action comment --url http://FQDN/rt"

The above statements will inform rt-mailgate to send all nagios notifications to the defined queue in RT. Check whether rt-mailgate works properly with the follow statement.

echo "checking functionality" | mail -s `rt-mailgatetesting' rt@FQDN

The above statement will generate ticket in RT with subject „rt-mailgate-testing. Create a queue with the same name in the configuration menu of RT. Also assign required rights to the users as well as groups. When evernagios will generate a notification, a ticket will be created in RT. The network monitoring system functionality can be tested by making any of the network node unavailable. This will generate a nagios notification which will create a ticket in RT. The ticket will be forwarded to all watchers in the defined queue of RT according to the priority.

VI. CONCLUSION

The paper addressed a hybrid approach which deals with a combination of agent and agentless network monitoring system by exploring four typical areas such as ease of deployment, adaptability, less network load and security aspects. Scenario of agent-based and agent-less monitoring has been brought out to clarify and support the hybrid concept. However, it is always a question of complying with quality requirements of network monitoring considering costs, budgets, time and network speeds that the choices are made in a typical industry scenario.

Abbreviations:
EG- Enterprise Group
HTTP- Hyper Text Transfer Protocol
HTML-Hyper Text Markup Language
SSH-Secure Shell
NetBIOS- Network Basic Input / Output System
RT- Real Time
UI- User Interface
HP- Hewlett Packard
TELNET- TELecommunication NETwork
GNU- GNU's Not Unix
SQL-Structured Query Language
URL- Universal Resource Locator
IIS- Internet Information Server

REFERENCES

[1] M. Schubert, A. Hay, D. Bennett et al., "Nagios3 Enterprise Network Monitoring," Designing Configurations for Large Organizations, Chap:2, pp.25-84, 2008.

[2] D. Oliveira, T. Vasques, F. Vieira, G. de Deus et al., "A management system for PLC networks using SNMP Protocol," presented at IEEE International Symposium on Power Line Communications and Its Applications (ISPLC), „10, Goias, Brazil, June 2010.

[3] A. Gomez, C. Dafonte, and B. Arcay, "3D Visualization for system and networks monitoring support," presented at 3rd IEEE Conference on Human System Interactions (HSI-10), A Coruna, Spain, July 2010.

[4] Rafiullah Khan, Sarmad Ullah Khan, Rifaqat Zaheer, and Muhammad Inayatullah Babar, "An Efficient Network Monitoring and Management System", International Journal of Information and Electronics Engineering, Vol. 3, No. 1, January 2013

[5]http://www.ijarcsse.com/docs/papers/Volume_3/9_ September2013/V3I9-0238.pdf, International Journal of Advanced Research in Computer Science and Software Engineering

[6] http://www8.hp.com/h20195/V2/GetPDF.aspx/ 4AA5-1319ENW.pdf

[7] Data communication and Networking, 4th Edition, Beherouz A Forouzan

[8]https://www.novell.com/coolsolutions/feature/16723. html

SECURING THE DATA CONFIDENTIALITY OF PATIENTS IN DISTRIBUTED M-HEALTHCARE CLOUD COMPUTING SYSTEM

Chandrika, Mahima.J, Yamuna.R, Swetha.P

[1,2,3] UG Student, [4]Assistant Professor Department of CS&E, RRCE, Visvesvaraya Technological University, INDIA.

Chandrika.cgowda@gmail.com, mahimajjayasimha@yahoo.com, yamunaranganath@gmail.com, shwetha6600@gmail.com

ABSRACT-Distributed m-healthcare computing system provides high quality patient treatment efficiently, but it brings number of challenges in safeguarding patients health information and identity privacy of patients. Many existing data access control and anonymous authentication systems are ineffective in distributed m-healthcare systems. To overcome this problem, in this paper, a authorized accessible privacy model (AAPM) is established. Distributed m-healthcare has three levels of security and privacy requirements and patients can authorize consultants by setting an access tree supporting flexible threshold predicates. Patients health information and personal information are verified by directly authorized, indirectly authorized and unauthorized physicians in medical consultation respectively satisfying the access tree with their own attributes.

Keywords- Access tree, distributed m-healthcare, privacy model, security, and threshold.

1. INTRODUCTION

1.1 Cloud Computing:

The cloud computing is based on internet computing that provides public processing resources and sources to computers and further devices on demand. It is a model for enabling universal on demand access to shared group of configurable computing resources. Cloud computation and storage solutions provide users and enterprises with various capabilities to store and process where data in third party Centre's. The cloud is a huge group of interconnected computers.

Computers can be personal computers or network servers; they may be public or private. Cloud of computers prolongs beyond a single company or enterprise. The application and data aided by the cloud are available to wide group of users, cross enterprise and cross stage. Access is via, the internet. Any authorized users can access these documents and applications from any computer over any internet connections. And to the users the technology and infrastructure behind the cloud is invisible. [1]

Cloud delivers a software platform that will enable customer to build an infrastructure-as-a service (Iaas) cloud. Cloud is built on the capabilities of existing virtualization management and physical servers provisioning solutions to deliver application to user that can be consumed in a self-service manner. Cloud optimizes the usage of physical and virtual organization through intelligent resource. Allocation policies and add the ability to flex applications elastically based on the demand.

1.2. m-healthcare:

M-healthcare is an abbreviation for mobile healthcare, a word used for the practice of medicines and public health records maintained by mobile devices. The term is most commonly used in reference to using communication devices like mobile phones, PDAs and tablet computers for health facilities and information, but also to affect emotional states. This paper contains Distributed m-healthcare cloud computing system provides high quality patient treatment for medical consultation by sharing patient's personal health information among hospitals. It brings series of challenges mainly how to ensure the protection and confidentiality of the patient's personal information and health information from the different type of attacks in the wireless communication channel. One of the main challenges is access control of patient personal health information. Access control is the policy driven limitation of access to system data and dialogs. Access permission defines whether a role or individual should have any access to all and, if so, exactly what the role or individual should be allowed to do the resource.[3]. Namely it is only the physicians or consultancies that have privilege to access the data that can recover the patient's personal health information during data sharing in distributed m-healthcare computing system.

Fig. 1 The Architecture for the m-healthcare system.

The basic m-healthcare system illustrated in Fig.1 mainly consists of three components: body area networks (BANs), wireless networks and the healthcare organizations furnished with their own cloud servers [1], [2]. The patient's personal health information is securely sent to the healthcare provider for the physicians to access and perform medical treatment. We further illuminate the unique characteristics of distributed m-healthcare cloud computing systems where all

the personal health information can be shared between patients suffering from the same disease for mutual support or among the authorized physicians in distributed healthcare organizations and medical research institutions for medical consultation.

2. WORKING METHODOLOGY

In m-healthcare social networks, the information about the person's health and his/her personal information are always shared between the patients located in respective social communities suffering from the same disease for mutual support, and across distributed healthcare providers (HPs) equipped with their own cloud servers for medical consultant.As to the security aspect, one of the main issues is access control of patients' personal health information, namely it is only the physicians or institutions that can recover the patients' personal health information during the data sharing in the distributed m-healthcare computing system.A fine-grained distributed data access control scheme is suggested using the technique of attribute based encryption (ABE). A rendezvous-based access control scheme provides access privilege if and only if the patient and the physician meet in the physical world. Now a days, a patient-centric and fine-grained data access control in multi-owner settings is constructed for safeguarding personal health data in cloud computing. However, it mainly focuses on the central cloud computing system which is not enough for efficiently processing the increasing volume of personal health information in m-healthcare computing system. Furthermore, it is not enough for to only promise the data confidentiality of the patient's health information in the cloud server model since the frequent communication between a patient and a physician can lead the opponent to conclude that the patient is suffering from a particular disease with a high possibility. Unfortunately, the problem of how to secure both the patients' data confidentiality and identity privacy in the distributed m-healthcare computing scenario under the malicious model was left untouched.To overcome this disdvantages, we propose a novel authorized accessible privacy model for distributed m-healthcare cloud computing system.

2.1. Authorized accessible Privacy Model (AAPM)

Several existing access control and anonymous authentication systems cannot be clearly exploited. To overcome the problem, in this paper, we established a novel that is authorized accessible privacy model(AAPM). Patients can authorize physicians by setting an access hierarchy supporting flexible threshold establishes to access their information provided to the hospital database. Then, based on it, by devising a new technique of attribute-based designated verifier signature, a patient self-controllable privacy-preserving authentication scheme realizing three levels of security and privacy requirement in m-healthcare computing system is proposed. The directly authorized physicians, the physicians who are indirectly authorized and the unauthorized persons in medical consultation can respectively decipher the personal health information and/or validate patients' identities by satisfying the access tree with their own attribute sets.

Lastly, the security proof and simulation results illustrate our scheme can resist different kinds of attacks and far outperforms the prior ones in terms of computational, communication and storage overhead. A basic architecture of a distributed m-healthcare cloud computing system is shown in Fig. 3

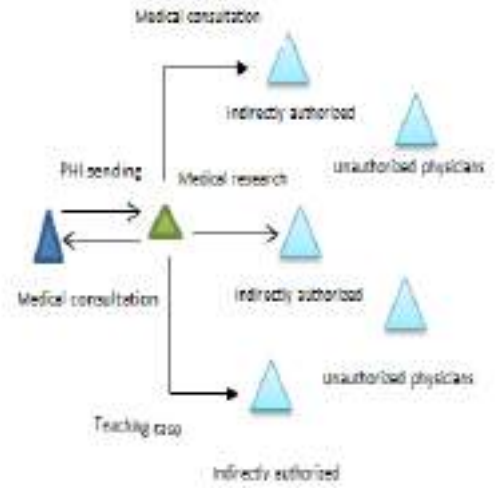


Fig. 2. Security and privacy levels in m-healthcare computing system.

There are three healthcare providers A; B;C and the research institution that is D, where Dr. Brown, Dr. Black, Dr. Green and Prof. White are functioning respectively. Each of them possesses their cloud server. It is assumed that patient P registers at hospital A, all her/his health information is stored in hospital database A's cloud server, and Dr. Brown is one of his directly authorized consultant. For medical consultation or other research purposes in cooperation with hospitals B;C and medical research foundation D, it is required for Dr. Brown to generate three indistinguishable transcript simulations of patient P's health information and share them among the distributed cloud servers of the hospitals B;C and medical research organization D.

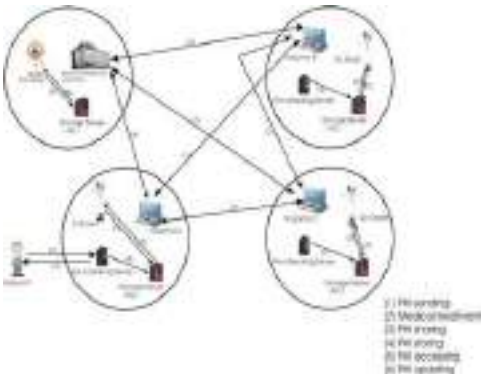


Fig. 3.An overview of m -healthcare cloud computing system.

Table.1 Access control on patient's data

Data	Directly authorized physicians	Indirectly authorized physicians	Unauthorized physicians
Personal Information	**YES**	**NO**	**NO**
Health Information	**YES**	**YES**	**NO**

3. CONCLUSION

In this paper, authorized accessible privacy model (AAPM) ,a novel is established and a patient self-controllable privacy-preserving authentication scheme realizing three levels of security and privacy requirement in the distributed m-healthcare system are proposed, followed by the formal security evidence and efficiency evaluations. Patients can provide privileges to the physicians by setting an access tree supporting flexible threshold predicates. The directly authorized doctors, the indirectly authorized doctors and the unauthorized physicians would know both personal data of patient's and the health information of patients, only the personal health information and nothing respectively. Finally, simulation results show our project far performs prior schemes in terms of storage, computational and communication overhead.

4. ACKNOLEDGEMENT

The author would like to thank Karnataka State Council for Science and Technology(KSCST) for funding this proposal (with serial number 39S_BE_1029). This work is carried out at project laboratory, department of computer science and engineering, RajaRajeswari College of Engineering

REFERENCES

[1]. Michael Miller *"Cloud Computing"*, Pearson Education, Inc, 2013.
[2]. Dr Kumar Saurabh *"Cloud Computing"* Wiley pvt, second edition,2013.
[3]. Raymond R. Ranko *"Corporate computer and Network Security"*, pearson Education, Inc, 2003
[4]. V. Goyal, O. Pandey, A. Sahai and B. Waters, Attribute-based Encryption for Finegrained Access Control of Encrypted Data, In ACM CCS'06, 2006.
[5]. M. Li, S. Yu, K. Ren and W. Lou, Securing Personal Health Records in Cloud Computing: Patient-centric and Fine-grained Data Access Control in Multi-owner Settings, SecureComm 2010, LNICST 50, pp.89-106, 2010.
[6]. J. Li, M.H. Au, W. Susilo, D. Xie and K. Ren, Attribute-based Signature and its Applications, In ASIACCS'10, 2010.
[7]. De-identified Health Information, http://aspe.hhs.gov/admnsimp/bannerps.htm.
[8]. J. Sun, X. Zhu, C. Zhang and Y. Fang, HCPP: Cryptography Based Secure HER System for Patient Privacy and Emergency Healthcare, ICDCS'11.
[9]. J. Zhou, Z. Cao, X. Dong, X. Lin and A. V. Vasilakos, Securing m-Healthcare Social Networks: Challenges, Countermeasures and Future Directions, IEEE Wireless Communications, vol. 20, No. 4, pp. 12-21, 2013

Three Party Authentication Using Quantum Key Distribution Protocol

Pratap M S, Drishya Nair, Ponnu Narayanan, Nitha Kamar, Aneesa C V
Department of Computer Science, Kvg College of engineering, sullia
pratapms17@gmail.com, drishyabhaskar25@gmail.com, ponnun7@gmail.com, nitharashad26@gmail.com, aneesa132@gmail.com

ABSTRACT-Cryptanalysis is an important branch in the study of cryptography, including both the Classical Cryptography and the Quantum one. In this paper, we analyze the security of Three-Party Quantum Key Distribution Protocol proposed recently, and point out that they are susceptible to a simple and effective Dense-Coding attack, through its architectural implementation in Classical Cryptography. The protocol focuses the operations of key generation and measurement in the trusted center's lab. With a symmetric key encryption method, the intruder is resisted from eliciting the confidential information transmitted between the legal users. It is shown that the eavesdropper Eve can totally obtain the session key by sending bits as the fake signal to sender and performing collective measurements after sender's encoding. The attack process is just like a dense-coding communication between Eve and sender, where a special measurement basis is employed. Furthermore, this attack does not introduce any errors to the transmitted information and consequently will not be discovered by sender and receiver. The attack strategy is described in detail and a proof of its correctness is given. Finally, the root of this insecurity and possible way to improve the protocol are discussed.

Keywords – encryption, decryption, Quantum Cryptography, QKDP

1. INTRODUCTION

As the most important application of system quantum cryptography, Quantum key distribution is used to produce and distribute a key, not to transmit any message data. This key can then be used with any chosen encryption algorithm to encrypt (and decrypt) a message, which can then be transmitted over a standard communication channel. It allows two users, generally called Alice and Bob, to privately share a random key through quantum key distribution protocol (QKDP) through which they obtain a shared session key through a trusted center. Hence the protocol is termed as Three-Party QKDP.

The existing scenario deals with the transmission of data directly from the sender to receiver where data communication software plays its intermediate role. It has been noticed so many issues concerning the privacy, such as dense-coding attack over the communication channel in transmitting large volume of personal and sensitive information. As the intruder encounters in the midst of the legal users in hacking the data, the communication is no longer reliable and secure. Three-party QKDPs proposed yet are susceptible to this attack in the sense that the third-party can totally acquire the session key which is being shared between the communicating users, by sending the entangled bits as the fake signal to the sender. Either of these users can never discover that eavesdropping has occurred since it won't introduce any error to the transmitted information. The protocol supports no measurement apparatus for the sender and receiver with the reason that they cannot take measures to detect eavesdropping. The attack process is just like a dense-coding communication between the third party and sender.

In order for a secure communication, prevention of dense-coding attack is necessary. To overcome these limitations of the existing system the new system is developed which in turn guarantees the legal users from eavesdropping.

Network Security is fast looming on the horizon as a potentially massive problem over the communication network. It covers a multitude of sins such as concerned with intruders trying to access remote services that they are not authorized to use or secretly modifying messages intended for other recipients. Many techniques are used to overcome the problems faced in communication and data transfer. In this paper we analyze the security of three-party quantum key distribution protocols proposed recently, which were susceptible to a the dense-coding attack.

Encryption is the standard method for making a communication private. Anyone wanting to send a private message to another user encrypts or enciphers the message before transmitting it. Only the intended recipient knows how to correctly decrypt or deciphers the message. Anyone who was eavesdropping on the communication would only see the encrypted message. Because they would not know how to decrypt it successfully and the message would make no sense to them. Cryptography uses this method in cryptosystems to maintain confidentiality of information. Since the security of most classical cryptosystems is based on the assumption of computational complexity, they might be susceptible to the strong ability of quantum computation and will become no longer secure once quantum computer appears. This in turn led to the advent of quantum cryptography. Different from its classical counterpart, quantum cryptography is the combination of quantum mechanics and cryptography, where the security is assured by physical principles which can stand against the threat from an attacker with the ability of quantum computation. QKD constitutes one of the branches of quantum cryptography. Shared keys are used for secure communication on insecure public network since the legitimate participant cannot ensure that the received session key is correct and cannot confirm the identity of the user. It focuses on the operations of key generation and its measurement in the trusted center's lab as the case here is

that the information is encoded in bits. So also they are insecure under Dense Coding Attack. Since cryptographic methods cannot be proven secure, Rijindael methods security rests on the fact that it is extremely difficult for an attacker to hack the key. With the implementation of a secure symmetric key encryption method named Rijindael algorithm, it is shown that eavesdropper is being prevented from obtaining the session key which is shared among the two communicating users.

2. METHODOLOGY

The system defines a secure Third-Party Quantum Key Distribution Protocol which is resistant to Dense-Coding Attack. In the protocol, the operations of bits generation and measurements are focused in the Trusted Center's lab. It verifies the correctness of the secret session key and authenticates the user to ensure that confidentiality is only possible for legitimate users. The sender encodes the session key into bits by performing parity checking. After the receive of key, decoding operation is being performed at the receiver terminal. With the implementation of a symmetric key encryption algorithm, the proposed QKDP detects Man-in-the-middle attacks on the data, thus offering a reliable communication between the legal users. User authentication and session key verification is being done in single step without any public discussion with the sender and receiver.
There are two types of Quantum Key Distribution Protocol, they are:

2.1 The Proposed 3AQKDP

This section describes the details of the 3AQKDP. Here, we assume that every participant shares a secret key with the TC in advance either by direct contact or by other ways.

2.2 The Proposed 3QKDPMA

The proposed explicit quantum key distribution protocol 3QKDPMA can be divided into two phases: the Setup Phase and the Key Distribution Phase. In the Setup Phase, users pre-share secret keys with the TC and agree to select polarization bases of qubits based on the pre-shared secret key. The Key Distribution Phase describes how the user and could share the session key with the assistance of TC and achieve the explicit user authentication.

3. RELATED WORK

Cryptography is the science of information security. Applications of cryptography include ATM cards, computer passwords, and electronic commerce. The combination of Implicit quantum key distribution protocol (3AQKDP) and explicit quantum key distribution protocol (3AQKDPMA) are used to form the new combination and demonstrate the following merits:
It will establish secure connection which can prevent attacks such as eavesdropping, man-in-the-middle and replay.
Reduction in communication rounds among existing QKDPs, improves efficiency of proposed protocols. A long Term secret key can be used and shared between two parties repeatedly.
Classical cryptography methods currently used are unsafe and cannot detect the existence of passive attacks such as eavesdropping. Hence the combination of both classical as well as quantum cryptography is proposed.
Research in authentication protocol has focused largely on developing and analyzing protocol that are secure against certain types of attacks. There is little and only scattered discussion on protocol efficiency. For each proven lower bound, an authentication protocol achieving the bound is also given, thus proving that the bound is a tight bound if the given optimal protocol is secure. Moreover, impossibility results of obtaining protocols that are simultaneously optimal with respect to the numbers of messages and rounds are given [6].

4. SYSTEM ANALYSIS

In Registration phase after providing username and password, user must generate one unique key for identification. That is Secrete key. This key will take part in our final key **(Quantum Key)**. At this instance our system will store every details such as username, password, secrete key, Registration Date and Registration Time.
After Registration for New User system redirects the user to Login Stage. At this stage the user must provide the relevant details which was noted or given through registration. The secret key generation is in separate class which will return.

4.1 Trusted Center Module

Trusted center module generates the key.it comprises of following modules:

4.1.1 Secret Key Verification

Verify the secret key received from the user and authenticate the user for secure transformation.

4.1.2 Session Key Generation

It is shared secret key which is used to for encryption and decryption. The size of session key is 8 bits. This session key is generated from pseudo random prime number and exponential value of random number.

4.1.3 Qubit Generation

To get secret key and random string, then convert into hex-code and then convert it into binary, find the least bit of two binary values and get the quantum bit of 0 and 1.

4.1.4 Quantum Key Generation

To generate the quantum key using the qubit and session key which depends on qubit combinations, such as :

1. If the value is 0 and 0, then 1/0.707(p[0]+p[1])
2. If the value is 1 and 0, then 1/0.707(p[0]-p[1])
3. If the value is 0 and 1, then p[0]
4. If the value is 1 and 1, then p[1]

4.1.5 Key Distribution

It distributes the original session key and qubit to the sender for encryption. Also, it distributes the qubit and the session key on the receiver side for decryption.

4.2 Receiver

Getting Authorization is the first stage in receive phase. If a user wants to receive a text from source user, he wants unique Identification. By using that Identification System, we can identify that that the person is an authorized person.
This phase or Receiver Module has Sub Modules. They are:
4.2.1. Registration
4.2.2. Login and
4.2.3. Receive Data

4.2.1 Registration

Registration is the Initial state for getting Authentication. By Providing username and Password user sets their Authentication. And System provides one more credentials that is Secrete key which is generated by the system for each user. By using **username**, **Password** and **Secrete key** system will identify the Authorized person. These values are stored in the Database quantum key in which **reg** table.

4.2.2 Login

If a user wants to send a file, he/she must log in by using his/her authentication credentials. In this module we have to give **username**, **password** and **Secret key** which was generated by the system.
If the user does not provide proper information or the given information is mismatched with database then our system shows Exception message immediately.
If the user's details are verified and matched with the existing database then our system allows the person to transmit the file.
After login the TCP program calls i.e. our Trusted Center program starts listen the client or sender. Through Login we send the sender's secrete key for Identification.

4.2.3 Receive Data

The main aim of this module is to decrypt a file. Decryption will happen only if the system gets a key from **Trusted Center (TC)**. So after verification of user identification system will send the current user's name and his/her secret key to **Trusted Center (TC).**
In Registration phase after providing username and password, user must generate one unique key for identification. That is Secret key. This key will take part in our final key **(Quantum Key)**. At this instance our system will store every detail such as username, password, secret key, Registration Date and Registration Time.
After Registration for New User system redirects the user to Login Stage. At this stage the user must provide the relevant details which was noted or given through registration.

5. SYSTEM MODEL

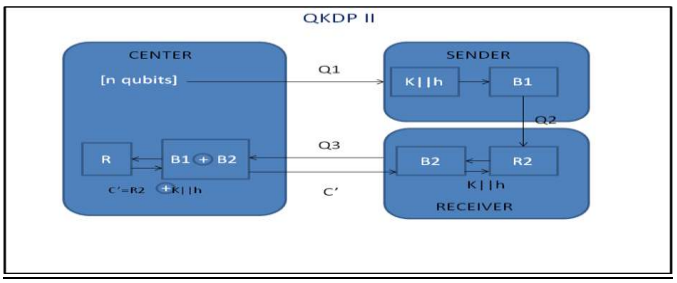


Fig 1: system model for QKDP

Algorithm for Logic Implementation

The Protocol is composed with the following steps:

1) The Trusted Center generates n bits and sends this sequence (Q1) to Sender (Alice).
2) After receiving Q1, Alice selects a u-bit random session key K and computes its m-bit hash value h=H (K) as the checksum, where u+m=n. Then Alice performs unitary operation U0=I on the i^{th} bit in Q1 if the i^{th} bit in K||h is 0. Furthermore, Alice generates an n-bit random string B1, and performs unitary operation U0=I on the i^{th} bit in Q1 if the i^{th} bit in B1 is 0. After these coding operations Alice sends the new sequence (denoted as Q2) to Receiver (Bob). Here

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3) After receiving Q2, Bob selects two n-bit random strings R2 and B2. Then he performs unitary operation U0 in each bit in Q2 according to R2, and then operation U0 on each bit according to B2. Afterwards Bob sends the new bits sequence (denoted as Q3) to the Trusted Center.
4) The Center informs Alice and Bob after the receiving of Q3.
5) Alice and Bob tell the Center B1 and B2 respectively.
6) According to B1⊕ B2, the Center recovers the original bases of bits by performing U0 on each bit as in step 2 and 3. Then the center measures all the bits in basis R= {0, 1}, obtaining the measurement results , C'=R2⊕(K||h). Finally, C will be announced to Bob by center.
7) Bob recovers K||h= R2 ⊕C' and verifies whether h=H(K). If it is correct, Bob obtains the session key K and tells Alice it is successful.

6. CONCLUSION

The fast streaming evolutions in the cyber world is undergoing drastic revolutions over the technologies and network environment. The demands are on the hike in the field of communication from Classical to Quantum networks. Yet the problem of utmost security over the confidential information and the Key remained as such. Key exchange is weak spot in many cryptosystems. Even the strongest cryptography is useless, if the key used to encrypt and decrypt the data are not secure. Mechanisms for key exchange exist yet are unsusceptible over powerful computers. QKDP II is tuned to cater this need, without putting additional overheads, affecting the network topology. The protocol implements a system assuring a reliably secure communication between the sender and receiver by analyzing the security of the existing QKDP which were susceptible to Dense-Coding Attack. The protocol focuses the operations of key generation and measurement in the trusted center's lab. With a symmetric key encryption method, the intruder is resisted from eliciting the shared

session keys between the legal users and thus used to encrypt and decrypt the message which in turn transmits over a standard communication channel. The Trusted Center constitutes a module that verifies the correctness of the secret session key and authenticates the user to ensure that confidentiality is only possible for legitimate users. It generates the random secret key from which Session key and dispenses it to the sender for encryption. Accordingly, announces the measure of bits to the receiver end by distributing it along with the session key for decryption. Generation of the Session Key is entrusted to the Sender module. Mean while, the Receiver module recovers the session key which is being formally generated by the Sender

.

7. FUTURE WORK

The current protocol has been chosen Classical Cryptography as its implementation domain. It will well suits if the same is applied on Quantum computers since the architectural design is constructed to cope with the Quantum Cryptographic scenario. Accordingly, improved bit rates can be achieved in exchanging secure keys. Any improvements in the hardware can favor for exploiting the opportunities at the application level. Quantum key distribution (QKD) systems have the advantage of being automatic, with greater reliability and lower operating cost than other secure networks.

The protocol is well fit to compete to future amendments at the software aspect so can be re-structured. Advanced Cryptographic algorithms can be implemented evaluating its security over credential information. If improved achievements are reflected, the benefits will be immense.

REFERENCES

[1] Dr Anand Rao, "Three party authentication key distribution protocol using implicit and explicit quantum cryptography , Vol. 2 No. 2 143-145, 2011.

[2] G. Li, "Efficient Network Authentication Protocols: Lower Bounds and Optimal Implementations," Distributed Computing, vol. 9, no. 3, pp. 131-145, 1995.

[3] A. Kehne, J. Schonwalder, and H. Langendorfer, "A Nonce-Based Protocol for Multiple Authentications," ACM Operating Systems Rev., vol. 26, no. 4, pp. 84-89, 1992.

[4] M. Bellare and P. Rogaway, "Provably Secure Session Key Distribution: The Three Party Case," Proc. 27th ACM Symp. Theory of Computing, pp. 57-66, 1995.

[5] J. Nam, S. Cho, S. Kim, and D. Won, "Simple and Efficient Group Key Agreement Based on Factoring," Proc. Int'l Conf.
Computational Science and Its Applications (ICCSA '04), pp. 645-654, 2004.

[6] H.A. Wen, T.F. Lee, and T. Hwang, "A Provably Secure Three- Party Password-Based Authenticated Key Exchange Protocol Using
Weil Pairing," IEE Proc. Comm., vol. 152, no. 2, pp. 138-143, 2005.

[7]. System Analysis and Design, Ellias M Award

[8]. C#.Net Programming Bible, Jeff Ferguson, Bian Patterson and Meeta Gupta

[9]. Software Engineering, Ian Somerville

[10]. SQL Server 7-The Complete Reference, Gayle Coffman

Reducing the Time Factor for analyzing the Errors and Fixing in Big Sensor data

Anusha G, Dr. N Guruprasad
Dept. of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru-560103.
anushagutta25@gmail.com, nguruprasad18@gmail.com

Abstract-- The big sensor data is so common in the industries and research application that grows so rapid that increase in volume and velocity is going to be meteoric. As the sensor data keeps on increasing it becomes a very difficult task to process using the on hand database tools and data processing applications. To overcome from the drawbacks the cloud computing has been provided with a unique feature of tackling with the challenges by providing the flexible massive storage, computing and software services in a lower cost. Many techniques has been developed for processing sensor data such as sensor cloud but it would not support fast error detection and location of error faster in big sensor data sets. In our proposed system real based approach for faster detection and location of errors in the sensor data sets. The error detection is based on scale free network topology which would dramatically accelerate the process. We fix the errors using the spatial and temporal correlation technique. This approach can reduce the time for detection's of the error and also location of the error in big data sets and main advantage is fixing the errors.

Keywords-- Big sensor data sets, Error detection, scale free network, types of errors, temporal correlation spatial correlation.

I. INTRODUCTION

The cause of data explosion in the present era the biggest challenge faces is processing of the big data. Since big data is collection of data sets and it so complex to process it as the data keeps on exploring. The traditional approach of human cognitive process which includes datasets which is beyond the ability to process the data in tolerable elapsed time which would be a major drawback since datasets keeps on accumulating day by day and becomes difficult task to process it. One of the major and important characteristic of big data is volume, velocity, value, veracity and variety. The big data sets can from any base such as meteorology, complex physics simulations, biological study and environmental research.

One important source of data set is collected by wireless sensor network (WSN). The WSN have feature of enhancing the ability of monitoring and interact with physical environment. Since there is corruption and lose of data due to presence of WSN in hardware inaccuracies in the node. It is necessary for data to be received clean and accurate. There is a need of effective detection and also cleaning of sensor big data is a major challenging and requires innovative solutions. WSN with the cloud can be called as complex network systems. As the complex network increases the data in accuracy and error has become an issue in real network application.

WSN big data error detection usually requires real time processing and also storage for massive sensor data which would also use the complex error model to detect the event of abnormality. In this paper we aim to develop a approach by having massive storage, scalability and also having computation power to detect error in big data sets from sensor data. The proposed error detection approach in this paper is by detecting the types of errors. The main work is to achieve time efficient approach in detecting the errors without compromising error detection accuracy and also the recovery of the error.

II. RELATED WORK

As there was increase in the data, the processing of the data sets also increased in era of data explosion. The different techniques for processing the data are as follows:

A) Big Data Processing on Cloud-- The cloud computing has a significant feature of big data storage and also interpreting it massive computation power. A design called "stream-as-you-go" is used for the processing the data has it is increased through the stream based management data architecture. Map Reduce is also used to process to analyze the incremental data and also process by dividing the data. But all this focus on the workload distribution, scalability and filtering of the data in speed but these approach is not enough for error detection and correction.

B) On- Cloud Processing for WSN -- The WSN are used in different fields such as in environment monitoring, military and also in scientific data collection. Since the variety of data has been collected from different fields through the WSN, there is a need of the sensor cloud for processing of the data. But due the nature of big data with the feature of volume, variety and velocity it is difficult to process the data in the complex network system. By effect of this using the cloud sensor it is difficult to develop the time efficient approach to detect the errors in the data sets and also to debug the complex network system in real time.

C) Data Error Detection in Sensor Network and Complex Network – The data error is unavoidable in real world complex network system. As there was dramatic increase in big data generated to locate the error was a quite challenging task with normal computing and network system.

Wang et al has proposed a classification of errors based on scenario analysis. It performed well in finding the errors detecting the errors. This work compares the robustness of four node network only, clustering co-efficient and centrality and hence can be extended for the complex network system.

Mukhopadhyay proposed error correction method for the WSN. It was performed on the intelligent sensor network. The intelligent sensor network correction of the error was faster. But as the big data sets increased its processing capability and time performance was extremely limited when the data was increasing.

In our approach we would detect the errors in the complex network by improve the time efficiency and faster detection and correction of the errors.

III. PROPOSED WORK

In this paper we discuss about reducing the time for analyzing the errors and fixing of the error in the big data sets. We would use Time Efficient Error Detection with would consist of Scale Free Sensor Network System (SFSNS) which is used to process unlimited amount of data and Model based Error detection (MBED). Before applying this approach we need to classify the types of errors. The errors can be classified based on the numeric data and text file data. Since the errors are so common in complex network system. The architecture of the proposed system is follows which clearly specify the structure and flow of the system which is been shown in fig 1.

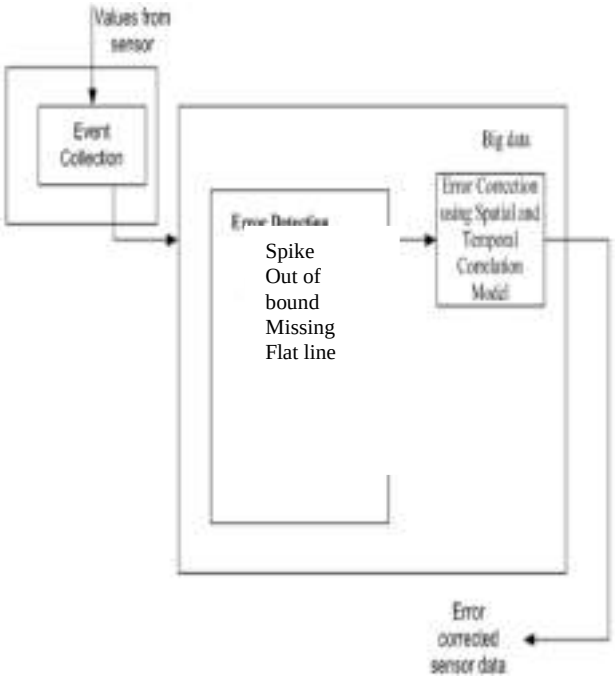


Fig-1 Architecture diagram

The errors are classified as follows

A) Types of Errors -- The flat line faults shown in fig-1 indicate that nodes in the network kept unchanged for unacceptable time series and for long duration. In real world application the transmitted data will have small changes over time flow.

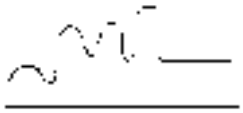


Fig-2 Flat line fault

The out of bound faults shown in fig-2, the value of the data can be observed based on the domain knowledge that is gained in general. In real world applications if a threshold is fixed for a wave, if any wave beyond that it is treated as out of bound error.

Fig-3 Out of bound fault

The data lost fault shown in fig-3, it means that there is missing of data over time during the data is generated and exchanged process which requires data cleaning.

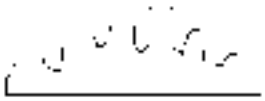


Fig-4 Data lost fault

The spike fault is shown in fig- 4 the spike error indicates that in a time series data items which would be out of prediction suddenly and normal over the time series.

Fig-5 Spike fault

B) SFSNS -- The WSN graph has a strong feature of scale free complex network such as cluster head WSN. The scale free network consists in-homogeneous and also only few nodes in graph have large number of links in the nodes. In reality the cluster head WSN is almost similar to the scale free network. The scale free network has a more clustered hierarchical nodes topology. By getting the hierarchical structure of the graph we can easily traverse the nodes and detect the error.

C) MBED -- The complex networks systems have a similar to the clustered network topology. When there is a data abnormality in the big data datasets, there two tasks to be completed for detection the errors. There are to detected abnormal data is the true error that is "false negative" for detecting a true error and "false positive" for the non-error data. Finally detecting and tracking of the

error source in which node. By these two tasks we can finalize successfully the error detection process. The error detection process needs to filter big data sets from the network. When there is data abnormality the whole network should be traverse for finalizing error source localization. In scale free network only few nodes in the hierarchy will have large set of links to the nodes. So based on node which has huge links can be grouped in the cluster so that the error can be located easily, so that we can need to navigate to search the error and the location of the source.

Finally for the error correction and for the recovery we use a different approach called spatial correlation and temporal correlations are used. In temporal correlation we can identify the sensors attack and the forged values are sent. In spatial correlation by having the knowledge of sensor we can correct the error easily. The proposed of clustering can reduce the time for detection of the error and also processing of the data of whole can be avoided.

IV. IMPLEMENTATION

In the implementation the sensor data is been collected through the simulation, the data sets through the simulation using the simulator. The sensor data is given to the pre processor to process the data. By using the error tracker I would track the errors. The algorithms used to track all four errors are mentioned below. The first used algorithm is for detecting the missing error algorithm.

Algorithm -1 Detection of missing error

Step 1 Collect the data sets that the id, time, value.
Step 2 Differentiate data sets for values based on "#".
Step 3 Initialize the sensor id=0.
Step 4 for each i=0 to i++.
Step 5 Find the value of diff=time[0]-time[1].
Step 6 for each i=2 to i++.
Step 7 Find the value of dn=time[i]-time[i-1].
Step 8 If dn= diff , no error.
Else missing error.

The detection of the flat line error algorithm is given in algorithm 2 below.

Algorithm-2 Detection of flat line error

Step 1 Initialize a variable time =0.
Step 2 Initialize obstime = -1.
Step 3 for each i=1 to i++.
Step 4 val= value[i]-value[i-1].
Step 5 if val=0.0
time++
obstime = 1
Step 6 Else time=0
Obstime = -1
Step 7 if time is greater than 5
Return flat line error

The detection of the spike error algorithm is given in algorithm 3 below.

Algorithm-3 Detection of spike error

Step 1 Get the value of the sensor.
Step 2 for each i=1 to i++
Step 3 Calculate prev=val[i-1]
Current=val[i]
next=val[i+1]
Step 4 if(current>prev &&current>next)
Avg=(prev+next)/2
Tim=Current/Avg
Step 5 if(tim>5)
Return spike error
Step 4 Else no error

The detection of the out of bound error algorithm is given in algorithm 4 below

Algorithm-4 Detection of out of bound error

Step 1 Initialize the value of maxvalue=60.
Step 2 for I=0 to i++
Step 3 if value is greater than max value
return out of bound error
Step 4 else return there is no out of bound error.

The entire above mentioned algorithm is implemented to detects the errors from the sensor data and finally corrected.

V. CONCLUSION

In the process of detecting the errors in the sensor network system, a new approach is been developed using the cloud computing. Initially the error classification in the sensor data sets is been differentiated. The correlation between sensor networks and the scale free networks is achieved. By using the correlation between both the error types and the scale free networks features a time efficient strategy for detecting the errors and also locating the errors in big data set is achieved. Finally the error detection approach can reduce the time for fast error detection in numeric big data sets and it achieve similar error selection ratio for non-scale free error detection. In accordance to the error detection of big data sets, it would also recovery correct the error and false positive ratio is also reduced.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Network", IEEE Comm. Mag., vol. 40, no. 8, 2002.
[2] Wireless Sensor Networks, Technology, Protocols and Application by K. Sohraby, D. Minoli and T. Znati.
[3] S. Mukhopadhyay, D. Panigrahi, and S. Dey, "Data Aware, Low Cost Error Correction for Wireless Sensor Network," proc. IEEE Wireless Comm. And Networking Conf.
[4] "Sensor Cloud," http://www.sensorcloud.com.
[5] S. Sakr, A.Liu, D. Batista, and M. Alomari, "A Survey of Large Scale Data Management

Approaches in Cloud Environment," IEEE Comm. Survey.

[6] M.C. Vuran and I. Akyildiz, "Spatial Correlation based on Collaborative Medium Access Control in Wireless Sensor Network"

[7] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big DataApplications," IEEE Trans. Parallel and Distributed Systems, 2013.

[8] Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. the ACM, vol. 53, no. 4,pp. 50-58, 2010.

[9] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud Computing and Emerging it Platforms: Vision, Hype, and Reality for Delivering Computing As the 5th Utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.

[10] L. Wang, J. Zhan, W. Shi, and Y. Liang, "In Cloud, Can Scientific Communities Benefit from the Economies of Scale?" IEEE Trans. Parallel and Distributed Systems, vol. 23,, no. 2, pp. 296-303, Feb.

Automating Infrastructure as a Code using Continuous Integration and Continuous Delivery

Bhanupriya H, Dr. Krishna A N, Ravi Shekhar Jha
PG student[1], Professor [2], DevOps Specialist [3] ,
Department of Computer Science, SJBIT, under VTU, Bangalore-60
bhanupriya15393@gmail.com, csehod@sjbit.edu.in, ravisjha@hpe.com

ABSTRACT-Every project team expects their project management to be automated and robust. With advancements in automation technologies, one can think of automating the manual built infrastructures like storage servers, project servers and deployments on those servers. We can achieve this automation of Infrastructure and build a continuous delivery pipeline which accelerates the speed of project development, testing and deployment using Emerging technologies. This paper depicts how team can leverage the Continuous Integration and Continuous Delivery concepts to develop their projects to increase the efficiency and speed up their releases.

Keywords - Configuration management, Continuous Delivery, Continuous Integration, GitHub, Jenkins

1.Introduction

Software Release In Organizations To Users Is Often Painful, Risky And Time Consuming. Continuous Integration And Continuous Delivery (Ci/Cd) Can Help Big Organizations Become As Slender, Agile And Innovative As Startups. Through Trustworthy, Low Risk Releases Ci/Cd Makes It Possible To Continuously Adapt Software Aligned With User Feedback, Market Shifts And Changes In Business Strategy [1].Test, Support, Development And Operations Work Together As One Delivery Team To Automate And Simplify The Build, Test And Release Process. Continuous Integration Is Outlined As Fundamental Practice In Extreme Programming Methodology, It Has Emerged As An Essential Element For Teams Doing Iterative And Incremental Software Delivery.Continuous Delivery Is Extension Of Continuous Integration, Which Ensures The Team That Every Change They Make To The System Will Be Releasable, And That We Can Release Any Version At Just Push Of A Button.

2. Problem Statement

Having a bad development workflow will be costly; it degrades the productivity of engineers in development to deployment cycles. A great workflow can make any good developer to be great and best ones to be exceptional. The most important problem that we face as software professionals is: If somebody thinks of a best idea, how do we deliver it to users as quickly as possible.

There are many software development methodologies which primarily focus on requirement management and its effect on the development effort. It is challenging to find what happens once requirements are identified, solutions formulated, developed and tested, how these activities joined together and synchronized to make process as efficient and reliable as team can make it? How do we

able developers, testers, and build and operation engineers to work together effectively?

The day of a software release tends to be a tense one. The process used to make the release of projects increases the degree of risk and terrifying sometimes. In many software projects, release is a manually intensive process; finally the application is started, piece by piece if it's a distributed or service oriented application. If any step is not perfectly executed, the application won't run properly. It is very difficult to identify what went wrong and where the error is.

2.1. Disadvantages of manual deployment:

The creation of extensive, detailed documentation that describes the steps to be taken and the ways in which the steps may go wrong during deployments. Confidence on manual testing to confirm that the application is running correctly. Repeated calls to the development team to explain why a deployment is going wrong on a release day. Frequent rectifications to the release process during the course of a release. Releases that take more than a few minutes to achieve. Releases that are unpredictable in their result to be rolled back.

2.2. Need for Automated testing & deployment

When deployments aren't fully automated, errors will occur every time they are executed. The question of interest is whether or not the errors are noteworthy. Even with excellent deployment tests, bugs can be hard to track down.

Automated deployments encourage cooperation, because everything is explicit in a script. Documentation has to make assumptions about the level of knowledge of the reader and in reality is usually written as a reference for the person performing the deployment, making it solid to others.

2.3. Risk in deploying to production like environment

Releasing into staging is the first time that operations team interact with the new release. In some organizations,

separate operations teams are used to deploy the software into staging and production. In this case, the first time an operations person sees the software is the day it is released into production. Either a production-like environment is costly enough that access to it is strictly controlled.

The development team assembles the correct installers, configuration files, database migrations, and deployment documentation to pass to the people who perform the actual deployment—all of it untested in an environment that looks like production or staging.

When the deployment to staging occurs, a team is assembled to accomplish it. Sometimes this team has all the necessary skills, but often in very large organizations the responsibilities for deployment are divided between several groups and it results in poor collaboration. It should also be possible to use the same automated process to roll back to a previous version of production if the deployment goes wrong.

3. PROPOSED METHOD

The Software release should be a low-risk, frequent, cheap, rapid, and predictable process. Our goal is to describe the use of deployment pipelines, combined with high levels of automation of both testing and deployment and comprehensive configuration management to deliver push-button software releases.

3.1. Every Change Should Trigger the Feedback Process

An operational software application can be usefully decomposed into four components: executable code, configuration, host environment, and data [2]. If any of them changes, it can lead to a change in the behavior of the application. Therefore we need to keep all four of these components under control and ensure that a change in any one of them is tested.

Executable code changes when a change is made to the source code. Every time a change is made to the source code, the resulting binary must be built and tested. In order to gain control over this process, building and testing the binary should be automated. Continuous integration is the practice of building and testing your application on every check-in.

This executable code should be the same operational code that is deployed into every environment, whether it is a testing environment or a production environment. Anything that changes between environments should be noted as configuration information. Any change to an application's configuration, in whichever environment, should be tested.

The tests will vary depending on the system, but they will usually include at least the following checks. The process of creating the executable code must work. This verifies that the syntax of your source code is valid. The software's unit tests must pass. This checks that your application's code behaves as expected. The software should fulfill certain quality criteria such as test coverage and other technology-specific metrics.

The software's functional acceptance tests must pass. This checks that your application conforms to its business acceptance criteria—that it delivers the business value that was intended. The software's nonfunctional tests must pass. This checks that the application performs sufficiently well in terms of capacity, availability, and security, and so on to meet its users' needs.

The software must go through empirical testing and a demonstration to the customer and a selection of users. This is typically done from a manual testing environment. In this part of the process, the product owner might decide that there are missing features, or we might find bugs that require fixing and automated tests that need creating to prevent regressions.

3.2. Open to feedbacks in early stages

The key to fast feedback is automation. With fully automated processes, your only limitation is the amount of hardware that you are able to throw at the problem. If you have manual processes, you are dependent on folks to get the job done. People take longer, they introduce errors, and they are not auditable. Moreover, performing manual build, test, and deployment processes is boring and repetitive—far from the best use of people. Developers should commit changes to their version control system frequently, and fragmented code into separate components as a way of managing large or distributed teams.

3.3. The Delivery Team must be reactive to Feedback

It is essential that everybody involved in the process of delivering software is involved in the feedback process. That includes developers, testers, operations staff, database administrators, infrastructure specialists, and managers.

A process based on continuous improvement is essential to the rapid delivery of eminent software. Iterative processes help establish a regular heartbeat for this kind of activity—at least once per iteration a retrospective meeting is held where everybody discusses how to improve the delivery process for the next iteration.

Finally, feedback is not noble unless it is acted upon. This requires discipline and planning. When something needs doing, it is the responsibility of the whole team to stop what they are doing and decide on a course of action. Only once this is done should the team carry on with their work.

3.4. Scaling of the process

One common complaint we hear is that the process we describe is idealistic. It may work in small teams, these critics say, but it can't possibly work in any huge, distributed project. Both the theory and the practice are as relevant to large teams as they are too small, and our experience has been that they work.

4. METHODOLOGY

4.1. Create a Repeatable, Reliable Process for Releasing Software

Releasing software should be easy. It should be easy because you have tested every single part of the release process hundreds of times already. It should be as modest as pressing a button. The repeatability and reliability derive from two principles: automate almost everything, and keep everything you need to build, deploy, test, and release your application in version control.

Deploying software eventually involves three things: 1) Provisioning and managing the environment in which your application will run (hardware configuration, software, infrastructure, and external services). 2) Installing the correct version of your application into it. 3) Configuring your application, including any data or state it requires.

The deployment of your application can be realized using a fully automated process from version control. Application configuration can also be a fully automated process, with the necessary scripts and state kept in version control or databases. Clearly, hardware cannot be kept in version control; but, particularly with the advent of cheap virtualization technology and tools like chef, ansible, the provisioning process can also be fully automated [3].

4.2. Automate Almost Everything

There are some things it is impossible to automate. Exploratory testing depend on experienced testers. In general, your build process should be automated up to the point where it needs specific human direction or decision making. This is also true of your deployment process and, in fact, our entire software release procedure.

Automation is a prerequisite for the deployment pipeline, because it is only through automation that we can guarantee that people will get what they need at the push of a button. However, you don't need to automate everything at once. You should start by looking at that part of your build, deploy, test, and release process that is currently the tailback. You can, and should, automate gradually over time.

4.3. Keep Everything in Version Control

Everything you need to build, deploy, test, and release your application should be kept in some form of versioned storage. All of the necessary stuff should be version-controlled, and the relevant version should be identifiable for any given build. That is, these change sets should have a single identifier, such as a build number or a version control change set number that references every piece.

It should be possible for a new team member to sit down at a new workstation, check out the project's revision control repository, and run a single command to build and deploy the application to any accessible environment, including the local development workstation. It should also be possible to see which build of your various applications is deployed into each of your environments, and which versions in version control these builds came from.

4.4. If it's risky test it more rather than at later stage

This is the most general principle on our list, and could perhaps best be described as a heuristic. Integration is often a very painful process. If this is true on your project, integrate every time somebody checks in, and do it from the start of the project.

If releasing software is painful, aim to release it every time somebody checks in a change that passes all the automated tests. If you can't release it to real users upon every change, release it to a production-like environment upon every check-in. If creating application documentation is painful, do it as you develop new features instead of leaving it to the end. Extreme programming is essentially the result of applying this heuristic to the software development process. Much of the advice in this book comes from our experience of applying the same principle to the process of releasing software [4].

4.5. Build Quality In

The earlier you catch defects, the cheaper they are to fix. Defects are fixed most cheaply if they are never checked in to version control in the first place. Firstly, testing is not a phase, and certainly not one to begin after the development phase. If testing is left to the end, it will be too late. There will be no time to fix the defects. Secondly, testing is also not the domain, purely or even principally, of testers. Everybody on the delivery team is responsible for the quality of the application all the time.

4.6. Done Means Released

For some agile delivery teams, "done" means released into production. This is the ideal situation for a software development project. However, it is not always practical to use this as a measure of done. The initial release of a software system can take a while before it is in a state where real external users are getting benefit from it. So we will choose the next best option and say that a functionality is "done" once it has been successfully showcased, that is, demonstrated to, and tried by,

representatives of the user community, from a production-like environment.

Start by getting everybody involved in the delivery process together from the start of a new project, and ensure that they have an opportunity to communicate on a recurrent regular basis. Once the barriers are down, this communication should occur continuously, but you may need to move towards that goal incrementally. This is one of the central principles of the DevOps movement [7].

4.7. Continuous Improvement

It is worth highlighting that the first release of an application is just the first stage in its life. All applications evolve, and more releases will follow. It is important that your delivery process also evolves with it.

The whole team should regularly gather together and hold a retrospective on the delivery process. Somebody should be nominated to own each idea and ensure that it is acted upon. Then, the next time that the team gathers, they should report back on what happened. This is known as the Deming cycle: plan, do, study, and act [8].

5. TOOLS USED

This section demonstrates how we can build Continuous Integration and Continuous Delivery pipeline and thus find the solution to the problem explained in 2nd section above. We need Version control systems (Source code Management), Build automation server (Jenkins), Configuration management tool (chef), Selenium (testing scripts), Servers for deployment at different environments (Development, Staging, Production) and any project code that has to be deployed residing in SCM (Source Code Management). We will design CI/CD (Continuous Integration and Continuous Delivery) pipeline after going through each of these selected tools.

5.1. Distributed Version control systems

The rise of distributed version control systems (DVCSs) is transforming the way teams cooperate. Where open source projects once emailed patches or posted them on forums, tools like Git make it extremely easy to pull patches back and forth between developers and teams and to branch and merge work streams. DVCSs allow you to work easily offline, commit changes locally, and rebase or defer them before pushing them to other users.

The core characteristics of a DVCS is that every repository contains the entire history of the project. GitHub has an additional layer of indirection: Changes to your local working copy must be checked in to your local repository before they can be pushed to other repositories, and updates from other repositories must be resolved with your local repository before you can update your working copy.

In the traditional model, committers acted as gatekeepers to the definitive repository for a project, accepting or rejecting patches from contributors. Due to the various above advantages we choose GitHub as version control system for building CI/CD pipeline.

5.2. Build automation server- Jenkins

Jenkins has a large pool of plugins allowing it to integrate with pretty much every tool in the build and deployment ecosystem. Once CI (Continuous Integration) tool jenkins is installed, it should be possible to get started in just a few minutes by telling your tool where to find your source control repository, what script to run in order to compile, if necessary, and run the automated commit tests for your application, and how to tell you if the last set of changes broke the software. It is general purpose job executer.

5.3. Configuration management tool-Chef

There are a number of solutions for managing operating system configuration, including which software and updates are installed, on an ongoing basis. Perhaps the most popular are ansible, Puppet, and Chef. We have used Chef as the tool for building CI/CD pipeline in this paper, it is a powerful automation platform that transforms complex infrastructure into code, bringing your servers and services to life[3].

Chef is built around simple concepts: achieving desired state, centralized modeling of IT infrastructure, and resource primitives that serve as building blocks. These very same concepts allow Chef to handle the most difficult infrastructure challenges on the planet. Anything that can run the chef-client can be managed by Chef. Chef recipes are the programs written in ruby used to achieve the desired configuration. Need to bootstrap the nodes/servers which has to be configured only the first time to establish connection between client and chef server. Figure1 shows the client server architecture of chef topology.

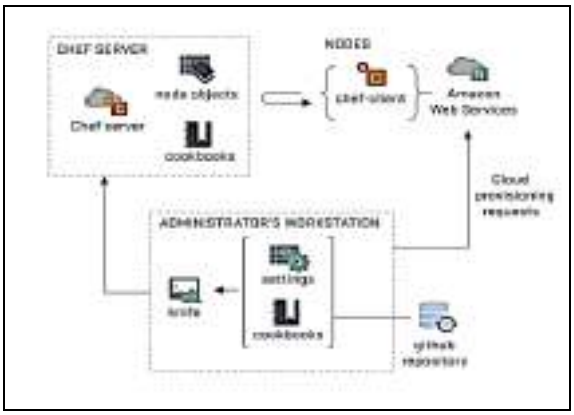


Figure 1: Chef Topology

5.4. Selenium scripts for testing

Selenium is a set of different software tools each with a different approach for supporting test automation. Most Selenium QA Engineers focus on the one or two tools that most meet the needs of their project, however learning all the tools will give you many different options for approaching different test automation problems.

5.5. Deployment servers

These are the servers used to deploy the code in different environments like Development where developers see how changes look like, Staging where testing can be done and Production where users are allowed to access once the application is released [12].

6.PROCESS

We can setup CI/CD pipeline using the above tools. For using Chef we need chef Development kit to program the necessary configurations. Chef Server has to be setup on Red Hat or any OS. It requires bootstrapping of nodes/servers (here its deployment servers) for the first time to establish connection between the server and those deployment servers which acts as client to the chef server. Configuration includes installation of necessary packages, setting environment variables to facilitate running of the application. Recipes are the program consisting of configurations necessary for deploying. It is uploaded as cookbooks, data bags, roles etc. on Chef Server which is later made to execute on Clients bootstrapped. These are applied to clients (servers) when we run the command "chef-client" on those servers. Thus Chef used for deployment on servers.

In Jenkins, jobs can be configured such a way that it pulls the application code that has to be deployed from the GitHub when a change is made to it and run the chef command "chef-client" on the server on which it has to be deployed i.e. development server. Job can also be configured/created such a way that deployment to staging can happen every night and tested every night each changes made and results are sent to the developers as a feedback in the form of an email. By this, we can configure multiple jobs to perform deployments on servers on different environment, testing of those applications and set to run on conditional basis.

Whenever the developer checks in the code to GitHub it creates a pull request and send a message to Jenkins to deploy code on development server so that developer can see the effect of changes he made to the application codebase. Build happens on development and notifies developer if deployment was successful on completion of executing "chef-client" on that server.

Every night the changes made to the application to be deployed in GitHub is deployed automatically to the staging server and selenium scripts are made to run on that staging environment to perform testing of the application. Results are sent via email to the developers as a feedback to rectify mistakes/errors if any. This is called Nightly build.

Once the development team feels the application is ready to release without any errors, code from GitHub can be deployed to production server where end users can access it. By this way, releases are made easy, less time consuming and robust. This can boost efficiency of development team and the business strategy.

Figure 2 shows the Deployment workflow cycle, Figure 3 shows the overview of how the pipeline can be setup. Figure 4 shows the feedback mechanism in CI/CD. Figure 4 shows the GitHub overview of how application code that has to be deployed looks like. Figure 6 is snapshot of chef server where we can manage cookbooks that is run on servers. Figure 7 shows the Jenkins setup that is necessary to setup automatic jobs to deploy automatically whenever necessary.

Figure 2: Workflow of Deployment Cycle

Figure 3: Continuous Integration and Continuous Delivery pipeline

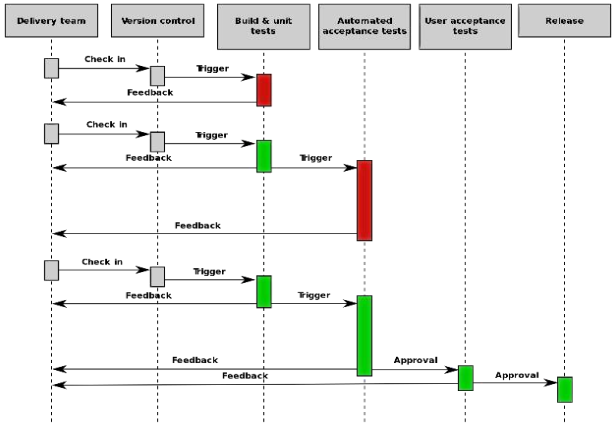


Figure 4: Feedback mechanism in CI/CD

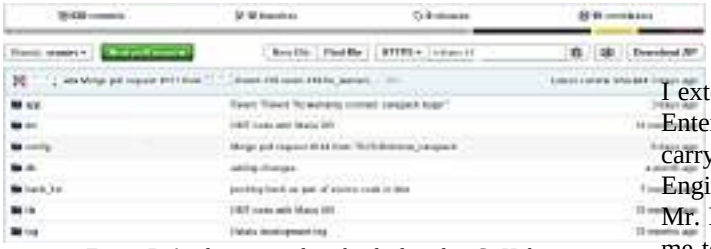


Figure 5: Application code to be deployed in GitHub

Figure 6: chef server

Figure 7: Jenkins job list and configurations

CONCLUSION

Management is vital to the success of every project. Good management creates processes enabling efficient delivery of software, while ensuring that risks are managed appropriately and regulatory regimes are complied with.

Finally, it is demonstrated that iterative delivery, combined with an automated process for building, deploying, testing, and releasing software exemplified in the deployment pipeline, is not only compatible with the goals of conformance and performance, but is the most effective way of achieving these goals. This process enables greater collaboration between those involved in delivering software, provides fast feedback so that bugs and unnecessary or poorly implemented features can be discovered quickly, and paves the route to reducing that vital metric, cycle time. This, in turn, means faster delivery of valuable, high-quality software, which leads to higher profitability with lower risk. Thus the goals of good governance are achieved.

ACKNOWLEDGEMENTS

I extend my sincere thanks to TSS R&D, Hewlett Packard Enterprise, Bangalore for giving me this opportunity to carry out this project. I am Grateful to Senior Software Engineering manager Mr. Aravind Desai and my mentor Mr. Ravi Shekhar Jha and other colleagues for supporting me to work on this project. I am also thankful to Head of Department Computer science, Dr. Krishna A N, SJBIT for his sincere support.

REFERENCES

[1] David Farley; Jez Humble "Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation, "Published by Addison-Wesley Professional, 2010.

[2] Beck, Kent, and Cynthia Andres, "Extreme Programming Explained: Embrace Change (2nd edition)", Addison-Wesley, 2004.

[3] Mischa Taylor, Seth Vargo "Learning Chef" Publisher: O'Reilly Media, Inc. Release Date: November 2014.

[4] George Spafford, Kevin Behr, Gene Kim, "The Phoenix Project" Publisher: IT Revolution Press, Release Date: January 2013.

[5] Donald Simpson, "Extending Jenkins" Publisher: Packt Publishing, Release Date: December 2015.

[6] Ben Straub, Chris Dawson, "Building Tools with GitHub", Publisher: O'Reilly Media, Inc. Published: February 2016.

[7] Adzic, Gojko, "Bridging the Communication Gap: Specification by Example and Agile Acceptance Testing", Neuri, 2009.

[8] Cohn, Mike, "Succeeding with Agile: Software Development Using Scrum", Addison-Wesley, 2009.

[9] Beck, Kent, and Cynthia Andres, "Extreme Programming Explained: Embrace Change (2nd edition)", Addison-Wesley, 2004.

[10] Duvall, Paul, Steve Matyas, and Andrew Glover, "Continuous Integration: Improving Software Quality and Reducing Risk", Addison-Wesley, 2007.

[11] ThoughtWorks, Inc., "The Thought Works Anthology: Essays on Software Technology and Innovation", The Pragmatic Programmers, 2008.
[12] Clark, Mike, "Pragmatic Project Automation: How to Build, Deploy, and Monitor Java Applications", The Pragmatic Programmers, 2004.

BIOGRAPHIES OF AUTHORS

Bhanupriya H, is currently studying 4th sem M.Tech (CSE) in SJBIT, Obtained Bachelor's degree under VTU. Presently working as an Intern, Hewlett Packard Enterprise, Bangalore as part of DevOps team. Has a publication in National Conference. Interested in Research in Automations.

Dr. Krishna A N is currently Professor & Head, Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru. He obtained his Bachelors and Master's degree in Computer Science and Engineering from University Visvesvaraya College of Engineering, Bangalore University, Bangalore and PhD degree from Visvesvaraya Technological University, Belagavi, Karnataka, India. He has publications in international conferences and journals. His research interests include image processing, pattern recognition and content-based image retrieval.

Ravi Shekhar Jha obtained Master's Degree in Computers and Technology from Mysore University, has work experience of 8.5 years. Presently working as DevOps Specialist, TSS R&D, Hewlett Packard Enterprise, Bangalore. He has worked extensively in different automation technologies and has a passion for automating processes as much as possible at every stage of software development lifecycle.

Privacy Preserving Mobile Access of Health Data Enabling Auditability

Deepashree N S, Tanushree k n
Department of Computer Science, R V College of Engineering, Bangalore-98, Department of Computer Science, Global Academy of Technology, Bangalore-98
deepashreens91@gmail.com, antanushree@gmail.com

ABSTRACT - Cloud computing is a new and virtually precise concept of computing technique, by which computer resources are shared dynamically through the Internet thus by appealing considerable and remarkable attention and interest from both academia and industry. This computing virtualization enables flexible and low cost computing thus enabling it outsource to the cloud servers thus making privacy a least concern. Although various schemes have been put forward to overcome the issue of privacy and safeguarding its information, but it seems natural that users might want to keep their identities secret and to review privilege control while they still get their privacy and so accessing this information should not cause reentrancy and an overhead during the communication. Hence, in this paper, we present a control on a semi-anonymous privilege scheme which ensures to address not only the privacy of the data but also the user identity privacy. Cipher-text policy decentralizes the central authority to limit the identity leakage and thus achieves semi -anonymity. The data is encrypted in two hierarchies one credential uses AES which encryption occurs at the local slot and one in the medium with server host, CPABE technique is used so to accomplish this task. In considering this entire scenario we can see the cipher-text generation can be done by protocols which results in thorough encryption which avoids the security breach thus making it semi anonymous to the respective attributes and thus enhancing the privileges to individual authority.

Keywords - Emergency medical technician, K-Anonymity, mHealth, Mobile Applications, Sever aided CPABE.

I. INTRODUCTION

Health care is the integral part of life in the human being. Health data accessing is slow evolvement and its procuring enables a well serviced health provisioning, enhances the quality of life and helps in reducing time for analysis and increase the fast treatment in medical emergencies anywhere-anytime. Benefits of cloud storage are easy access of the data to one's awareness anyplace, anytime, anyhow [2]. The proposed cloud-assisted mobile health networking is motivated by the power, flexibility, convenience, and cost efficiency of the cloud-based data/computation outsourcing paradigm. This paper introduces the private cloud security which can be considered as a service offered to mobile users. The proposed solutions are built on the service model. A software as a service (SaaS) provider provides private cloud services by using infrastructure of the public cloud providers (e.g., Amazon, Microsoft, yahoo, Google). An efficient encryption technique can be used for secure access to and storage of data on public cloud server, moving and searching encrypted data through communication channels while protecting data confidentiality [1]. Mobile devices (e.g. smartphone, PDA and laptop) have become the primary computing platform for many users because of their mobility and network connectivity [3]. Mobile users outsource the data processing tasks to the private cloud which stores the processed results on the public cloud. The cloud-based service model supports the implementation of privacy mechanisms since intensive computation and storage can be shifted to the cloud, leaving mobile users with the lightweight tasks. Pros do not merely help in diversifying the technology thus making analysts to keep a look on the challenges. The cloud emphasizes or helps us to analyze our proposed system thorough scenario and thus would provide a basement for the invention of new algorithms which is really what the technology needs. The main entities involved in our system are illustrated in Fig. 1. The system involves in user collecting their health data through the health tracking patches, electrocardiogram sensors. Emergency medical technician (EMT) is a physician who performs the emergency treatment. By user and EMT refer to the person and the associated computing facilities. The computing facilities are mainly mobile devices that carried around such as smart-phone, tablet, or personal digital assistant. Each user is associated with one particular private cloud. Multiple private clouds are supported on the same physical server. Private clouds are always available to handle health data on behalf of the users because private clouds are always online.

However the above environment does not suit the requirements of the methodology. At this present scenario small and medium scaled organizations cannot afford to build up an own cloud environment to use the fundamentals of identity. In this scene there is much less effort has been made during those interactive protocols. User's identities, which are reported with their attributes, are opened to key generators, and the generators issue private keys according to their attributes. But it appear natural that users might want to keep their identities secret while they still get their private and accessing this information should not cause reentrancy and an overhead during the communication.

In considering this entire scenario we can see the cipher-text generation can be done by protocols which results best with the key generation tactics to avoid the security breach.

This can be very advantageous in the situations like medical emergencies. The private cloud will process the data to add security before it is stored on the public cloud. Public cloud is the cloud infrastructure owned by the cloud providers like Amazon and Google which offers massive storage and rich computational resource. We assume that at the bootstrap phase, there is a secure way between the user and his/her private cloud, e.g., secure home Wi-Fi network, to obtain a long-term shared-key. After the bootstrap phase, user will send health data over insecure network to the private cloud residing via the Internet backbone. Nowadays, physicians are increasingly utilizing mobile health (mHealth) applications in clinical care [6].

II. RELATED WORK

According to Shamir et al In the IBE, the sender of a message can define an identity such that only a receiver with exactly identical identity can decrypt it. This is totally a sound variation from Public-key Encryption. However this method provides good resiliency but compromises if the technology development is known. To mitigate this IBE – Fuzzy Identity-Based Encryption which is also synonymously known as Attribute-Based Encryption (ABE) is introduced. In their work, an identity is observed as a set of descriptive attributes. Different from the IBE, where the decryption could decrypt the message if and only if his/her identity is exactly the same as what specified by the encryption, this fuzzy IBE enables the decryption in which there are 'identity overlaps' exceeding a pre-set threshold between the one specified by encryption and decision of encryption policy is made by different parties.

Personal Health Record (PHR) service is an emerging model for health information exchange [4-5]. Mobile devices help in reducing this clutter, such as home care and remote monitoring enable the people in their flexible lifestyle and cause minimal interruption to their daily activities. In addition, it significantly reduces the hospital occupation, allowing patients with higher need of in-hospital treatment to be admitted. Fine! All these scenarios are possible but people admit to realize that they would completely lose their personal information and identity once it activates in the cyber space. This take place around because in a survey 8 million patients' health information was leaked over a couple of years. But why this medical data should be kept private rather than allowing somebody to have a research on it. Of course there are some quite good reasons for it. An employer may not find convenient to hire someone with certain diseases. A mutual fund or brokerage insurance firm may refuse to provide features once they know about the history of the disease of the patient. Despite the paramount importance, privacy problems are not addressed adequately at the technical level and efforts to keep health data secure have continually fallen short. This is because protecting privacy in the cyberspace is significantly more challenging. Thus, there is an important need for the development of viable protocols, architectures, and systems assuring privacy and security to protect sensitive and personal digital information.

III. PROPOSED SYSTEM

A. AES Encryption

The encryption process made up of the combination of various classical techniques like substitution, rearrangement and transformation encoding techniques. The modifications include addition of an arithmetic operation and a route transposition cipher in the attacks iterative rounds. The encryption and decryption modules in this algorithm include the Key Expansion module which generates Key for all iterations The Key expansion module is extended to double the number of iterative processing rounds in order to increase its exception against unauthorized attacks.

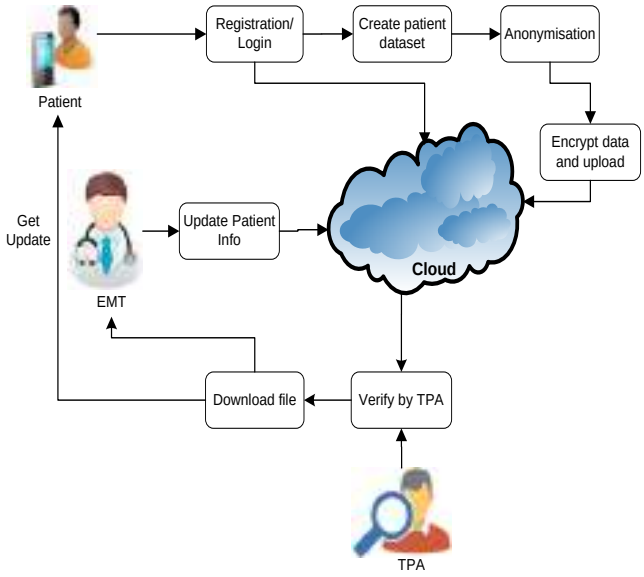


Fig. 1: Block diagram of Proposed System

Advanced Encryption Standard (AES) algorithm is not only for security but also for great speed. Both hardware and software implementation are faster still and replaces DES. AES encrypts data blocks of 128 bits in 10, 12 and 14 round depending on key size as explained above can be implemented on various platforms especially in small devices. It is carefully tested for many security applications.

Sub Bytes: The first transformation, Sub Bytes, is used at the encryption site. To substitute a byte, we interpret the byte as two hexadecimal digits.

Shift Rows: In the encryption, the transformation is called Shift Rows.

Mix Columns: The Mix Columns transformation operates at the column level; it transforms each column of the state to a new column.

Add Round Key: Add Round Key precedes one column at a time. Add Round Key adds a round key word with each state column matrix; the operation in Add Round Key is

matrix addition. The last step consists of XO Ring the output of the previous three steps with four words from the key schedule. And the last round for encryption does not involve the "Mix columns" step.

It is very important to know that the cipher input bytes are mapped onto the state bytes in the order $a_{0,0}, a_{1,0}, a_{2,0}, a_{3,0}, a_{0,1}, a_{1,1}, a_{2,1}, a_{3,1}$ and the bytes of the cipher key are mapped onto the array in the order $k_{0,0}, k_{1,0}, k_{2,0}, k_{3,0}, k_{0,1}, k_{1,1}, k_{2,1}, k_{3,1}$. At the end of the cipher operation, the cipher output is extracted from the state by taking the state bytes in the same order. AES uses a variable number of rounds, which are fixed: A key of size 128 has 10 rounds. A key of size 192 has 12 rounds. A key of size 256 has 14 rounds.

It is very important to know that the cipher input bytes are mapped onto the state bytes in the order $a_{0,0}, a_{1,0}, a_{2,0}, a_{3,0}, a_{0,1}, a_{1,1}, a_{2,1}, a_{3,1}$ and the bytes of the cipher key are mapped onto the array in the order$k_{0,0}, k_{1,0}, k_{2,0}, k_{3,0}, k_{0,1}, k_{1,1}, k_{2,1}, k_{3,1}$. At the end of the cipher operation, the cipher output is extracted from the state by taking the state bytes in the same order. AES uses a variable number of rounds, which are fixed: A key of size 128 has 10 rounds. A key of size 192 has 12 rounds. A key of size 256 has 14 rounds.

Decryption: Decryption involves reversing all the steps taken in encryption using inverse functions like Inverse shift rows, Inverse substitute bytes, Add round key, and Inverse mix columns. The third step consists of XOR-ing the output of the intermediates

B. SERVER AIDED-CPABE

On the other hand, CP-ABE has a solution to all these problems and thus solves partially the overhead involved. In the CP-ABE, cipher-texts are created with an access structure, which specifies the encryption policy, and private keys are generated strictly based on the users attributes. A user can access the cipher-text only if his attributes in the private key and the access tree specified in the cipher-text match. By doing so, the encrypted holds the ultimate authority about the encryption policy. Also, the already issued private keys will never be modified unless the whole system crashes and the system's master key are lost.

There is a model called multi-authority system, where each user has an ID and they can interact with each key generator (authority) using different pseudonyms this technique finds no replications to the method above. One user's different pseudonyms are tied to his private key, but key generators never know about the private keys, and thus they are not able to link multiple pseudonyms belonging to the same user. In fact they are even not able to distinguish the same user in different transactions. Also, the whole attributes set is divided into N disjoint sets and managed by N attributes authorities. That is, an attribute authority will only issue key components which it is in charge of. In this setting, even if an authority successfully guesses a user's ID, it knows only parts of the user's attributes, which are not enough to figure out the user's identity. In addition, many similar literature works have been published to create more advanced schemes where data needs to be securely and efficiently protected, which in turn served as the base of the research on security protocol in cloud computing environment. This attributes about the CP-ABE. However, we additionally tag it with the server for the second way of encryption which is called Server aided CP-ABE which happens at the host side providing along the security and the access privileges for the unit. However the detailed steps which associates with SA-CPABE is illustrated with the six steps below.

Setup: is identified with the small expression. The setup point issues two arguments one is the security parameter and the other one is total attribute description and thus showcasing the output with the public parameters PP (the data to be encrypted) and the master secret key MSK.

$$(\lambda, U) \rightarrow PP, SK \quad (1)$$

Pre-compute: The pre-computation algorithm takes as input the public parameters PP and outputs a temporal key TK and an intermediate cipher-textIC. The user keeps TK locally, and stores IC on its storage server to save local storage resources.

$$(PP) \rightarrow IC, TK \quad (3)$$

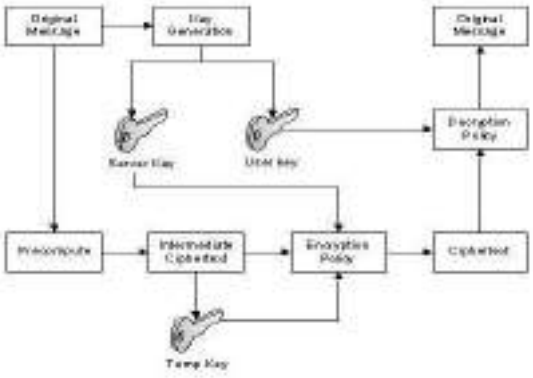


Fig. 1: Block diagram of the SA-CBABE

Encrypt: In the encryption side, three arguments are passed to the function as inputs an intermediate cipher-text IC, a temporal key TK, a message M, and an access structure A thus producing an intermediate encrypted output CT.

$$(IC, TK, M, A) \rightarrow CT \quad (4)$$

Transform: The cipher-text transformation algorithm takes as input a server key K_{server} for attribute set S and a ciphertext CT that was encrypted under A. It outputs the partially decrypted cipher-text CT if $S \in A$ and the error symbol $\perp$ otherwise.

$$(K_{server}, CT) \rightarrow CT \quad (5)$$

Decrypt: The decryption algorithm takes as input a user private key $K_{user}$for S and a partially decrypted ciphertext CT that was originally encrypted under$A$. It outputs the message M if $S \in A$ and the error symbol $\perp$ otherwise.

$$(CT, K_{user}) \rightarrow M \quad (6)$$

C. ANONYMITY

Health data service is a common and popular approach for attracting wide application users. K-anonymity is an

important measure for privacy to avoid the disclosure of personal data [7]. Software as a Service is one of the best and preferable methods in cloud computing that can be implemented by cooperation of various services and provides real-time services through the network.

The main idea is to provide secure and also anonymous online services of medical data among cloud computing infrastructure in specific organizations. Security can be enhanced in many ways like access control, anonymity, cryptography protocols and etc although there is a tradeoff between security enhancement level and system performance. Since Security implications should be applied thoroughly and specifically thus imposing to heavy burden on system processes. In all these cases we see that securing identity of an individual is primary task and selecting on how many attributes we need to perform is to be chosen based on the requirement and the criterion.

The k-anonymity model was first described in the context of data table releases. In this section we reiterate their definition and then proceed to analyze the merits and shortcomings of k-anonymity as a privacy model. The k-anonymity model distinguishes three entities: individuals, whose privacy needs to be protected; the database owner, who controls a table in which each row describes exactly one individual; and the attacker. The k-anonymity model makes two major assumptions: The database owner is able to separate the columns of the table into a set of quasi-identifiers, which are attributes that may appear in external tables the database owner does not control, and set private columns, the values of which need to be protected.

The term referred as two sets as public attributes and private attributes, respectively. Secondly the attacker has full knowledge of the public attribute values of individuals, and no knowledge of their private data. The attacker only perform linking attacks' linking attack is executed by taking external tables containing the identities of individual, and some or all of the public attributes that appear in a row of a table released by the database owner then we say that the individual is linked to that row. Specifically the individual is linked to the private attribute values that appear in that row. A linking attack will succeed if the attacker is able to match the identity of an individual against the value of a private attribute. As accepted in other privacy models (e.g., cryptography), it is assumed that the domain of the data and the algorithms used for anonymization are known to the attacker. Ignoring this assumption amounts to "security by obscurity," which would considerably weaken the model. The assumption reflects the fact that knowledge about the nature of the domain is usually public and in any case of a different nature than specific knowledge about individuals. For instance, knowing that every person has a height between zero and three meters is different than knowing the height of a given individual. Under the k-anonymity model, the database owner retains the k-anonymity of individuals if none of them can be linked with fewer than k rows in a released table. This is achieved by making certain that in any table released by the owner there are at least k rows with the same combination of values in the public attributes. Since that would not necessarily hold for every table, most of the work under the k-anonymity model focuses on methods of suppressing, altering, and eliminating attribute values in order that the changed table qualify as k-ananymous

IV. Results and discussions

The fig 3 described below invokes the interface of the proposed system which supports various features which include the navigational ability for home, patient, emt, tpa.

Fig. 3: Block Diagram of the proposed system

The next scenario would be to login with the credentials or else use the sign up if you are a new user the sign up form is described with its specific attributes shown in the fig 4. on successful completion we get a onetime password which is shown in fig 5.

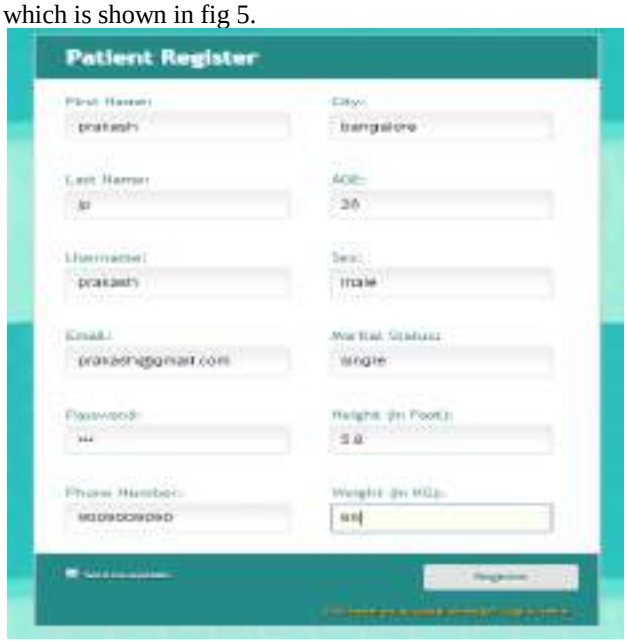


Fig. 4: Patient register

Fig. 5: one time password verification

The next process would be to use a partial encryption using AES encryption to modify the scenario of the data into unreadable format and the file thus obtained is fed to the next phase of encryption called CPABE algorithm.

Fig.6: File uploads system

Fig.7: files upload and anonymity setting

The complete encrypted file thus obtained using the encryption standard would display an unreadable format like this thus enhancing the security and showing its potentiality

Fig. 8: Encrypted data

Thus the entire above scenario indicated describes about the way the working and backend scenario of the algorithm looks like. Now if the person wants to get his data. The key should be available with him. Thus in the fig 9 we can see the owner and the candidate key along with file to be decrypted.

Fig. 9: Decryption of the data

The final interface describes about the attributes of the person but not its identity thus satisfying the criterion of anonymity.

Fig.10: Anonymity interface

V. CONCLUSION

This paper described an approach called cloud assisted mobile access and pointed out their strengths and limitations. This paper tells about the protection of the medical details and its anonymity in cloud. The proposed system builds privacy into mobile health systems with the help of the private cloud and provides a solution for privacy-preserving data storage by integrating a CP-ABE based key management for unlink ability. The system also investigated techniques that provide access control (in both normal and emergency cases) and audit ability of the authorized parties to prevent misbehavior, by combining anonymity controlled threshold signing with advanced encryption standard encryption. As future work, we plan to devise mechanisms that can detect whether users' health data have been illegally distributed, and identify possible source(s) of leakage (i.e., the authorized party that did it).

Acknowledgements

I express my immense gratitude to my guide Dr.S.R.Swamy, Professor, Department of Computer Science and Engineering, R V College of Engineering, Bangalore, for his estimated guidance and valuable suggestions in bringing out this work successfully

REFERENCES

[1] Hasan Omar Al-Sakran , "Accessing Secured Data In Cloud Computing Environment " International Journal of Network Security & Its Applications (IJNSA) Vol.7, No.1, January 2015.

[2] P. Umaeswari"Achieving Secure Data Access in Cloud Computing" Middle-East Journal of Scientific Research 23 (Sensing, Signal Processing and Security): 363-369, 2015.

[3] Yu Jin, Chuan Tian Heng He, Fan Wang," A Secure and Lightweight Data Access Control Scheme for Mobile Cloud Computing" 2015 IEEE Fifth International Conference on Big Data and Cloud Computing

[4] Raseena M ,Harikrishnan G R, "Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Broadcast Encryption" International Journal of Computer Applications (0975 8887)Volume 102 - No. 16, September 2014

[5] Chang-Ji Wang, Xi-Lei Xu, Dong-Yuan Shi and Wen-Long Lin "An Efficient Cloud-based Personal Health Records System Using Attribute-Based Encryption and Anonymous Multi-Receiver Identity-Based Encryption" Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing 2014.

[6] Vassiliki Koufi1, Flora Malamateniou1, George Vassilacopoulos. "Privacy-Preserving Mobile Access to Personal Health Records through Google's Android" International Conference on Wireless Mobile Communication and Healthcare - "Transforming healthcare through innovations in mobile and wireless technologies" (MOBIHEALTH)

[7] Khuong Vu and Rong Zheng JieGao" Efficient Algorithms for K-Anonymous Location Privacy in Participator Sensing".

Biographies and Photographs

Deepashree N S received Bachelor's degree in Information Science and Engineering from Sri Bhagwan Mahaveer Jain College of engineering Visveswaraya Technological University Belgavi in 2012. Currently pursuing MTech degree in Computer science and engineering in RV college of Engineering, Visveswaraya Technological University Belgavi. My research interests include Cloud computing security and network security and member of Computer Society of India (CSI)

Tanushree K N received Bachelor's degree in Computer Science and Engineering from Government Engineering College Visveswaraya Technological University Belgavi in 2011. Received Master degree in Computer Science and Engineering from East West Institute of Technology Visveswaraya Technological University Belgavi in 2014 Currently Working as Assistant professor in the Department of Computer science and engineering in Global Academy of Technology Visveswaraya Technological University Belgavi. My research interests include Cloud computing security and network security and member of Computer Society of India (CSI)

Intelligent Software Agent based Image Fusion in Wireless Multimedia Sensor Network

Prof.Dr.P.K.Srimani, Bhanu K. N.
Former Chairman, Dept. of Computer Science and Mathematics, Bangalore University,
Bengaluru, Karnataka, India
Associate Professor, Dept. of MCA, Rajarajeswari College of Engineering, Bengaluru, Karnataka, India
bhanu.kn@gmail.com, profsrimanipk@gmail.com

ABSTRACT-Wireless Multimedia sensor networks (WMSNs) can be used in monitoring disaster management, military operations, agriculture, building structures, etc. Multi sensor image fusion has been of great interest and challenge in WMSNs. This paper proposes an Intelligent agent based context aware image fusion scheme for Wireless Multimedia Sensor Networks. The proposed scheme operates as follows: Every sensor node and sink node deployed is running with agent platform. Sensor node agency resides in each of the sensor node and sink agency in sink node. The proposed scheme uses BDI (Belief, desire intention) model for in-network image fusion in WMSN. Sensor node agency performs context sensing and context interpretation by using BDI model and the interpreted context is sent to the sink node. Sink node agency receives the context information and generates the fusing agents (mobile), which are responsible for fusion of images from active sensor nodes. Mobile agent roams around the network, visits all the active sensor nodes and fuses the image. Mobile agent returns the fused image to sink node. The scheme is simulated to test the operation effectiveness.

Keywords - Wireless Multimedia Sensor Networks, Context-Aware computing, Software Agents, Cognitive Agents, BDI-architecture, Image Fusion.

I. INTRODUCTION

Wireless Multimedia Sensor Networks (WMSNs) comprises of tiny sensor nodes that have capability of sensing, computation, and wireless communications which are camera enabled. Recent development in the sensor technology has led to the development of low cost tiny sensor nodes enabled with CMOS cameras, which are combined with distributed signal processing and multimedia source coding techniques that are economically feasible. A tiny sensor node communicates within a short distance and collaboratively works to fulfill the application specific objectives of WMSNs [1] and [2]. Large number of sensors can be deployed for many applications that require unattended operations. These sensors are capable to communicate either to their neighbouring nodes or directly to an external Base Station (BS), called as sink. The deployment of more number of sensors allows sensing over larger geo-graphical regions with greater accuracy. Using conventional methods of data gathering and processing of multimedia data in WMSNs could lead to some of the problems like energy consumption, redundant data transmission, increased latency, bandwidth overheads, etc. The inclusion of context-awareness in WMSNs can solve these problems to a greater extent. Con-text can be identity, activity, location or time. Context is usually used to represent the information type in a system. To conserve the network life time of the WMSN, context aware computing and software agent technology paradigm together can give optimal solutions.

Usage of context-aware computing [3] in various applications helps to retrieve the information quickly. According to [4], a system is context-aware if it uses context to facilitate the relevant information and service to the user. Context aware information gathering in WSNs needs to have suitable information and measure of the context, which used to represent the system. Context-Aware computing mainly helps to get the relevant information from the environment, which in turn saves the energy consumption. WMSNs can be used for monitoring in various applications such as agriculture, disaster areas, health care, military, buildings, forests, animals, industrial control, etc. From recent literature it is found that WSN have great challenge to monitor the militant activities in the battle field. WMSN's can be used for various applications in military such as, army movement monitoring, ammunition monitoring, regulating friendly troops, etc. The sensors must be equipped with various visual aids, so that they can generate some really interesting and very useful data.

Intelligent Software agents can be employed for informa-tion fusion to prolong the network life time by eliminating the redundant information. Agents can be static or mobile. Static agents reside at particular location and perform tasks autonomously either by interacting with user or other agents at environment in the network and perform autonomous tasks by collaborating with other mobile/static agents or users. Mobile Agent based applications mainly fall in the areas like: network management, electronic commerce, wireless multimedia

sensors, grid computing and grid services, distributed data mining, multimedia, human tracking, security, etc.

Some of the related works are as follows. The military requirements for flexible wireless sensor networks have been provided in [5]. It describes the evolution of military sensor networking devices by identifying three generations of sensors along with their capabilities and also presents some of the existing developer solutions. The work presented in [6] investigates the design trade-offs for using WSN for implementing a system, which is capable of detecting and tracking military targets such as tanks and vehicles. The system estimates and tracks the target based on the spatial differences of the target object signal strength detected by the sensors at different locations. In [7] [8] a survey of mobile agent based applications is presented.

The work presented in [9] considers the tradeoff between the increase in the data aggregation required to reduce the energy consumption and the need to maximize the information integrity. A position-based aggregation node election protocol for wireless sensor networks, where aggregation node election is done to support asynchronous sensor network applications [10]. A load balancing data gathering algorithm that forms different groups of sensor nodes is described in [11]. A technique to extend the WSN operational time by organizing the sensors into a maximal number of disjoint set covers that are activated successively is presented in [12]. Active sensors are responsible for monitoring events and for transmitting the collected data, while nodes from all other sets are in a low-energy sleep mode.

The work given in [13] presents a method for fusing of the sequences of images obtained from multimodal surveillance cameras and subjected to distortions typically for WSNs. The scheme uses the Structural Similarity Measure (SSM) to measure a level of noise in regions of a received image in order to optimize the selection of regions in the fused image. Dual-Tree Complex Wavelet transform (DT-CWT) is used in the algorithm for region-based image fusion to fuse the selected regions. SOAR [14] is an excellent example for automated flight control and battlefield simulation which is developed by using cognitive agent based systems. The work mainly describes the military application scenario, where there is no predefined knowledge.

Our contributions are as follows: (1) Employing software agents at the sensor nodes to facilitate the cognition capabilities, which interpret and deliver context aware information in reliable way, (2) Intelligent decision making based on context, (3) Wavelet based image fusion code is embedded in mobile agent for image fusion, and (4) Fused image transmission reduces bandwidth requirement.

The rest of the paper is organized as follows. Proposed cognitive agent based image fusion in military sensor networks is discussed in section II. Simulation model is presented in section III. Results are analyzed in section IV. Finally, section V concludes the paper.

II. SOFTWARE AGENT BASED IMAGE FUSION

This section provides the preliminaries, network environment, and complete description of the proposed work.

A. Preliminaries

This section provides definitions of some terms used in describing the proposed work.

J. Sink Node: Sink node is responsible for collecting information and coordinating overall processing of the information, It has its own knowledge base which is called as Sink Black Board (SBB). SBB has all the information about the sensor node such as, node id, geographical location, communication range, bandwidth required etc. It mainly generates the fusion agent, which is required for fusion.

Critical information: An image sensed by the sensor node which is a critical object (such as gun, enemy movement, enemy vehicles, etc. in military). Based on the sensed image importance, context will be interpreted as critical and the sink node is informed.

Non-critical information: This relates to the less critical information such as, lighting conditions, fog, temperature, etc. Such information may be fused on-demand by the sink node.

Emergency context (critical context): Whenever the sensed image matches with any one of the critical images in the database of a sensor node, then we define it as an emergency context. This information will be sent to the sink node, which triggers fusion process with a single level wavelet fusion code.

Simple context: If the sensed image does not match with any one of the critical images in the knowledge base of sensor, then it is considered as simple context. Once simple context has been interpreted; information will be sent to sink node which triggers fusion process with two level wavelet fusion code.

Belief set: It is the belief set generated by sensor node based on the sensed parameters ψ_1, ψ_2.....ψ_N and actions taken.

Beliefs: It is the database comprising of belief sets generated by sensor node.

Image signal strength: It is entropy difference between the two considered images.

Image correlation: It is the degree of similarity between the images.

JJ. Network environment

Network environment is shown in figure 1. It comprises of heterogeneous sensor nodes and a sink node. Sensor nodes are geographically distributed over the network area, which collects the data periodically. Sensor node comprises of an agent platform with static and mobile agents, camera and other sensory devices such as light condition, temperature, humidity,

etc. The nodes sense data and send the sensed information to the sink node using wireless multihop communication. We assume that all the nodes in the network (sensor nodes and sink node) are static and have some initial energy. During deployment phase, all senor nodes have same energy. It is assumed that sensor nodes have capability to reconfigure the transmission power. All the sensor nodes are equipped with Global Positioning System (GPS), processor and transceiver for communication. Each sensor node communicates with neighboring sensor nodes within its communication range. Senor nodes participate in aggregation (active nodes) only if the sensed values in a particular time window drift by a given threshold.

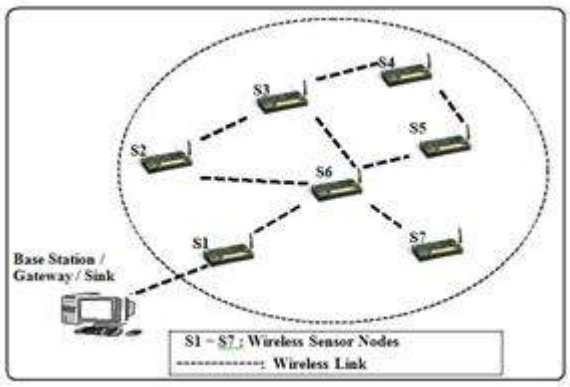


Fig.1. Network environment

C. Proposed work

Each agency employs a set of static and mobile agents to perform dedicated tasks. Sensor node agency performs context sensing and context interpretation based on the sensed image and sensing time. It comprises of node manager agent and context agent. Sink node agency initiates fusion process and it comprises of sink manager agent and fusing agent.

Agents use the Belief, Desire and Intention (BDI) architecture, which facilitate the cognition capabilities to make the intelligent decisions.

The usage of BDI architecture for agents presents the following benefits: WSN's exhibit dynamic changes in the environment and their parameters too. BDI agents have the capability of quickly adapting to such environment; the sensing capabilities and the sensing parameters may be many and can keep changing, therefore the beliefs regarding the environment can be regularly updated by using BDI architecture; autonomous decisions can be made based on the criticalness of sensed parameter.

The belief generation and action selection is done as follows:

• In the context of the BDI-frame work for the proposed work, here we will use AgentSpeak (L) to model the BDI system [15]. We will follow all the notations and expressions of the AgentSpeak (L) agent. An AgentSpeak (L) agent consists of the belief set and plan clause.

• In general the belief set {BN} can be formalized as BN = {ψS1N, ψS2N,.....,ψSNN , CN}. Where CN is criticalness of information which is either 1 or 0. NMA calculates the image signal strength of the sensed image out of all other sensed parameters. Agent that monitors the sensor node takes the decision CN. CN =1 indicates change in the entropy (critical) whereas CN=0 means no change in the entropy (non-critical).

• Based on these belief sets B1, B2 ...BN, the Node Man- ager Agent's (NMA) plan clause can be given as:

$$\text{goal} : \{B1 \quad B2, ..., \quad BN \leftarrow C_1....C_N\} \quad (1)$$

where B_N is the belief and C_N is an action.

Basically, AgentSpeak model works in the following steps.

[5] The agent selects an event that has occurred based on sensor status and images.

[6] The agent generates the plans with matching conditions, where the matching factor is defined by the monitoring agent that resides in node or the user.

[7] Among all the plans, agent identifies the plan with satisfying preconditions.

[8] The plan is then added to the intention stack. This intention stack is executed by popping out topmost plan and performing the first C_I, ifC_I is event.

Let us say, Ψ = MA plan set, $\Psi_I = i^{th}$ plan clause, C = goal which triggerss Ψ_I, BS= the belief set { B1^B2…….^Bn} We use Ψ(BS, C) to denote the body of ψI and also assume that the plan consists of number of action plans.

$$\Psi I(BS , C) = \begin{cases} \{(C1(P), C2(P)), \text{head}(\Psi I(BS , C) \\ (C1(P))), \text{tail}(\Psi I(BS , C)(C2(P)))\} \end{cases} \quad (2)$$

Where each action plan (CI(P)) is a well-defined action performed by the MA i.e, if action selected is C1 = critical information then MA generates Critical Context Interpreter Agent (CCIA) and if action selected is C2 = non-critical information then NMA generates Non-Critical Context Interpreter Agent (NCCIA). Both CCIA and NCCIA perform the actions as mentioned in below subsections.

Critical Context Interpretation Agent (CCIA): It is a static agent generated by NMA for interpreting the context. CCIA gets the sensed image from NMA and compares the sensed image with the images in the NBB. A histogram is used for the comparison of the images as given in context interpretation mediate value between high and low histogram value for a gray scale image) by the NMA. If the value after comparison of the images is below the threshold, then, the context is emergency type (i.e., image sensed is nearly same as that in the data base), else if the value is above threshold, then the context is interpreted as simple context (i.e., image sensed is nearly same as that in the data base). In case, where emergency context is detected, the corresponding information will be sent to the sink node. Sink node initiates fusion process using 1 level wavelet fusion code embedded in a mobile agent (called fusing agent). If simple context is detected sink

node initiates fusion process with 2 level wavelet fusion code embedded in a mobile agent.

Non-Critical Context Interpretation Agent (NCCIA): It is a mobile agent which is meant for storage and fusion of the non-critical information. If NMA decides in favor of the non-critical information, which mainly relates to fog, temp and other environmental conditions. Then this information will be stored by the agent itself and updates NBB. If suppose user at the sink node requires the non-critical information, user communicates to NMA of sensor node for this information. NMA triggers NCCIA. Once the agent is triggered, it visits all other active nodes (collects the information present in the KB/NBB) and fuses only the non-critical information from active nodes and returns to sensor node. The information will be communicated to sink node. This helps in conserving the energy as it will deliver the non-critical information only as and when user wants it. It also helps the user to have an overall aggregated information about the environment.

Fusing Agent (FA): It is a mobile agent equipped with image fusion code (1 level and 2 level wavelet fusion code) that migrates from one active node to another active node (we assume the agent itinerary to be given by SMA) depending on the routing information provided by the SMA. The agent visits an active node, fuses the image, and moves to another active node along with the fused image. The agent will use correlation model, which is kept inside the sensor node to fin d the correlation of data/images. If the value is high it means that the data between the two nodes is highly correlated, then it will classify it as the fusion node and fuses the data, else it will classify it as non-fusion node and moves on to the next node until it visits all active nodes.

D. Agent interactions

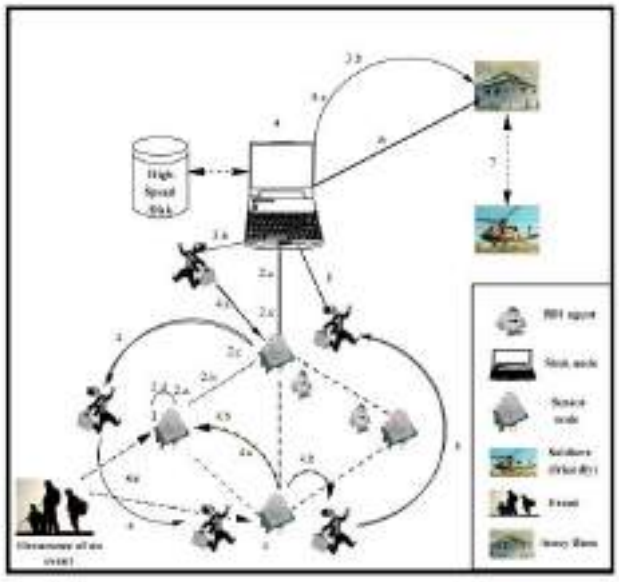


Fig. 2. Agent interaction

Figure 2 presents the agent interaction sequence of the proposed scheme:

1) Upon occurrence of an event in an environment. Sensing agent gathers the information through sensors about the target in the environment.

2) It updates the Knowledge Base (KB) and Belief sets of NMA. NMA decides that the sensed parameters are critical information; hence it generates CCIA, which in-turn starts interpreting the context, viz., emergency context and simple context.

   a) Histogram is used to interpret the context by fixing some threshold level, if the histogram level is below the threshold, then the context is considered as the emergency context and the node will send this information to the sink node.
   b) On the other hand if the value of the histogram is above the threshold, then the context is considered as simple context and this information is sent to the sink node.
   c) if NMA decides that the sensed parameters are non-critical information, then it will generate NC-CIA, which stores this information and also in-turn updates the KB/NBB. if user wants the informa-tion, then NCCIA visits all active nodes and fuses non-critical information.
   d) if the sensed parameter is of the environmental conditions, then the BDI agent configures the cam-era lens of the node accordingly as per lighting conditions.

3) After getting the context information or the image directly from the sensor node, the sink node will perform two functions:
   a) If the sink node receives emergency context information, then it generates FA with 1 level wavelet fusion code and sends it for information fusion.
   b) If the context is simple context, then the sink node generates FA with 2 level wavelet fusion code and sends it for fusion of information

4) FA visits the sensor node and collects the data and fuses the data. Then it will move on to next node (as per the routing table). When FA reaches the second node, it classifies the nodes as fusion and non-fusion nodes by triggering a process called as the correlation calculation process, again we divide this into two steps:
   a) FA when in the second node, calculates the correlation of the data with the first node.
   b) if the correlation value is high, then the FA will consider it as the fusing node and fuses the data else it will go to next active node.

5) FA will migrate to the sink carrying the fused informa-tion and updates the SBB (Sink Black Board).

6) Finally, the sink node will send the fused information to the control room for further actions.

III. SIMULATION

We have simulated proposed scheme for various network scenarios using C programming language with a confidence interval of 95%. In this section simulation model, performance parameters and result analysis are presented.

A. Network model

Wireless senor network is generated in an area of l × b square meters. It consists of N number of static nodes, placed randomly. Each node is associated with energy E_F joules, transmission range R meters. The communication environment is assumed to be contention-free. The transmission of packets is assumed to occur in discrete time. A node receives all packets heading to it during receiving interval unless the sender node is in non-active state. We assumed the channel as error free. Sensor MAC protocol (S-MAC) [16] is used for media access. Free space propagation model is used with propagation constant β.

Propagation model: Free space propagation model is used with propagation constant β. Transmission range of WMSN node communication radius is r for a single hop distance d meters. It is assumed that at any given time, the value of transmitted power is NP_{ow} milliwatts for every node.

Battery model: Image sensor nodes are deployed in the battle field, recharging of the nodes at the target is difficul t. So, we have considered a solar cell recharging model [Xiaoming Fan et al., 2004] and a layered clustering model to deal with the restrict energy consumption under the consideration of visual quality. The system lifetime can be prolonged by rechargeable solar cell that can be recharged by solar panel in daytime. Image sensor nodes consumes $node - batt$ millivolts to sense an image.

B. Simulation Procedure

To illustrate some results of the simulation, we have taken A = 100, B =200 meters and N=1 to 5, num=5 to 15, node − batt=90 millivolts, Th= 50%, 60 %, 70 %). Gray scale image of varying size rows × columns= (32 × 32, 64 × 64, 128 × 128, 256 × 256, (8, 12, 16, 24) bits/pixel, Present Signal strength= (30%,50%, 60%,70%, netBW = 4MBP S, Propagation Constant β =3.5, and F code = (4 , 8, 12) Kbytes.

Begin

- Generate the WMSN for the given radius and number nodes.
- Sense the parameters and generate the active nodes
- Apply the proposed context aware fusion model.
- Compute the performance of the system.

End

The following performance parameters are assessed:

Fig.4. Sensed image 2

Fig. 5. Fused image

Fusiontime (in sec) Vs No. of active nodes

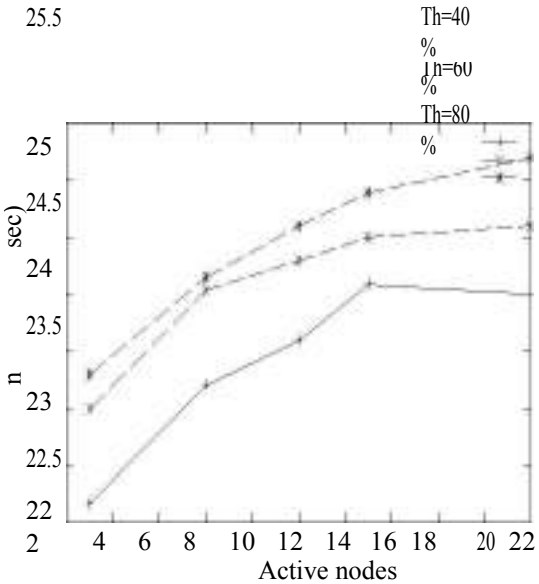


Fig. 6. Fusion time Vs. Active nodes

Figures 3 and 4 shows the sensed images (we have taken first and last images of the sensor nodes, when there were a total of 12 active nodes) and In the simulation,

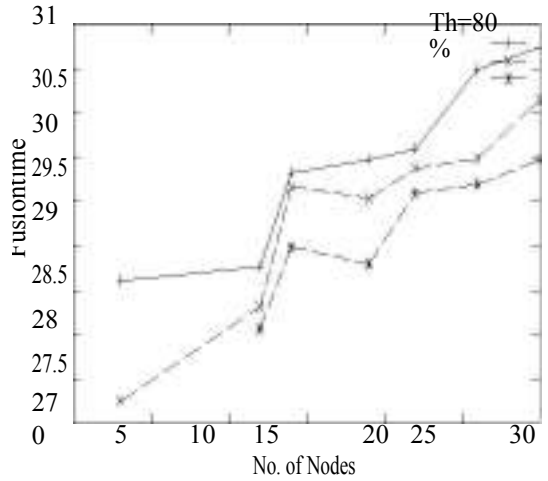


Fig. 7. Fusion time Vs. Number of Nodes

Utilized bandwidth Vs. No. Of Nodes

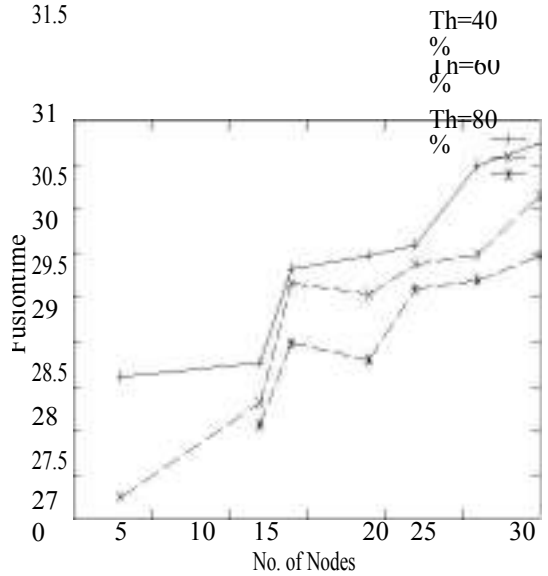


Fig. 8. Utilized bandwidth Vs. Active nodes

are taken from fusion process of the last active node. Figure 5 is the fused image.

Figure 6 depicts the analysis of fusion time. It shows that as the number of active nodes increase, the fusion time increases, but there is no abrupt increase in the fusion time even if there is increase in number of active nodes. Figure 7 shows that as number of deployed nodes increase along with increase in threshold, number of active nodes decrease and thus fusion time decreases accordingly.

In figure 8, different gray scale images are taken and their used bandwidth against number of active nodes has been analyzed. Bandwidth used increases as there is increase in the number of active nodes for each gray scale

image.

IV. CONCLUSION

This paper presented an intelligent software agent based con-text aware image fusion in WMSNs to form an infrastructure for image fusion. In an environment where source nodes are close to each other, and considerable redundancy exists in the sensed data, the source nodes generate a large amount of data, which not only wastes the scarce wireless bandwidth, but also consumes a lot of battery energy. BDI based intelligent agent has been used to interpret the context, and the given frame-work can be extended for various sensor input parameters.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey", Journal of Computer Networks, vo l. 38, no. 4, 2002.

[2] I.F. Akyildiz, Melodia. T.and Chowddhury K.R, "A Survey on Wireless Multimedia Sensor Networks", Journal of Computer Networks, v ol. 51, no. 921-960, 2009.

[3] Kirankumar B. Balavalad, S. S. Manvi and A. V. Sutagundar, "Context Aware Computing in Wireless Sensor Networks", In proceeding s of International Conference on Advances in Recent Technologies in Com-munication and computing, 2009.

[4] A. K. Dey, and G. A Bowd, "Towards A Better Understanding of Context and Context Awareness", In proceedings of 1st internationa l symposium on handled and ubiquitous computing, Karlsruhe, Germany, pp. 304-307, 2002.

[5] Michael Winkler, Graeme Barclay, "Theoretical and pract ical aspects of military wireless sensor networks", Journal of Telecommunica tions and Information Technology (JTIT), February, 2008.

[6] Tatiana Bokareva, Wen Hu, Salil Kanhere, " Wireless Sens or Networks for Battlefield Surveillance", Land Warfare Conference, Br isbane, 2006.

[7] Abdelkader Outtagarts, "Mobile Agent based applicatio ns: a survey", International Journal of Computer Science and Network Security, vol.9 no.11, 2009.

[8] Laura, G, Sergio, P., Andrew, T., "Efficient data aggrega tion in wireless sensor networks: an entropy-driven analysis", In Proceedi ngs of IEEE 19th International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC 2008, pp. 1 - 6, 15 - 18 Sept. 2008.

[9] Levente, B., Pter, S.," Position-Based Aggregator Node Election in Wireless Sensor Networks", International Journal of Distr ibuted Sensor Networks, Hindawi Publishing Corporation, 2010.

[10] Mihalela, C., Ding-Zho,D., "Improving Wireless Sensor Network Life-time through Power Aware Organization", Springer Journal of Wireless Networks, 11, pp. 333-340, 2005.

[11] Jianming, Z., Xiaodong, HU.," Improved Algorithm for Min imum Data Aggregation Time Problem in Wireless Sensor Networks", Jour nal of Systems Science and Complexity vol. 21, no. 4, pp. 626-636, 2008.

[12] Xin, G., Lin, G., Xin, G., Wang, Tomoaki, O., "A new load balancing and data collection algorithm for energy saving in wireless sensor networks", Springer Journal of Telecommunication Systems (JT S), pp. 313-322, 2010.

[13] Dejan Drajic, Nedeljko Cvejic, "Adaptive Fusion of Mul timodal Surveil-lance Image Sequences in Visual Sensor Networks", IEEE Trans actions on Consumer Electronics, vol. 53, no. 4, 2007.

[14] J Ones, R. M., L Aird, J. E., N Ielsen, P. E., C Oulter, K. J., K Enny, P., AND Koss, F. V. "Automated intelligent pilots for co mbat flight simulation", AI Magazine vol. 20, no. 1, pp. 2741, 1999.

[15] A. S. Rao. "AgentSpeak(L): BDI agents speak out in a logi cal com-putable language", Technical Note 64, Australian Artificia l Intelligence Institute, pp. 42 - 55, 1996, Also appeared in Agents Breaking Away (LNAI 1038).

[16] W. Ye, J. Heidemann, D. Estrin, "Medium Access Control Wi th Coor-dinated Adaptive Sleeping for Wireless Sensor Networks", I EEE/ACM Transactions on Networking, Vol. 12, No. 3, pp.493 - 506, 2004.

Dr. P.K. Srimani, F.N.A. Sc., is a Former Chairman, Dept. of Computer Science and Mathematics, Bangalore University, Bengaluru. She has more than thirty years of teaching and research experience in the field and has successfully guided more than a dozen Ph.D. students.

Bhanu K N is currently working as Associate Professor in the Department of Computer Applications, Rajarajeswari College of Engineering. She obtained her MCA, from IGNOU, NewDelhi, M.Phil., in Computer Science from VMU, Salem & presently pursuing Ph.D., in Computer Science from Rayalaseema University, AP.

EAMC: using Fog Computing and RRP for Exigency Alert Service

Deepika M, Srinivasa R
[1]PG Scholar, [2]Associate professor, Department of Computer Science, RRCE, VTU, Bengaluru, INDIA
deepikamgowda4@gmail.com , srinivasa.r@rediffmail.com

ABSTRACT-Timing is the most important factors in exigency management. Exigency notification techniques must be trouble free and quick, in order to have efficient response for any disaster, health-fix, act of terrorism, etc. In this paper, we present service architecture for exigency alert, using Fog computing cloud computing and Reliable Routing Protocol (RRP). Fog computing brings cloud resources close to the underlying devices and is used for offloading resource constrained devices and Reliable Routing Protocol (RRP) is used to maximize the reliability of data collection and control command delivery in wireless sensor networks. A smart phone based service, known as Exigency Alert Mobile Cloud (EAMC) provides a quick way of notifying the predicate exigency-handling department, applying the services of Fog and RRP for offloading as well as pre-processing purposes. The service sends the location of incident and contacts to the appropriate exigency-handling department and to family members automatically through already stored contact numbers. RRP aims to discover multiple bidirectional routes between a sensor node and a sink node. RRP achieves load balance by sending data packets via the route with lighter workload. The emergency related information is synchronized automatically from Fog to the Cloud.

Keywords-- Fog computing, Low end-to-end delay, Micro data center (MDC), Mobile cloud computing, Reliable data packet delivery.

I. INTRODUCTION

Exigency refers to an immediate risk posed to life, health, environment, or asset. Exigency has no calendar and one can come across any sort of such situation anywhere, at any time. Whether in the form of an accident, terrorism, robbery, kidnapping, fire breaking out, building collapse, murder. It has become urgent to have a simple and quick way to notifying the concerned department to deal with the disaster. Exigency management depends a lot on how quickly and exactly the situation is informed to the right people. With today's age of advancements in smart phone technology, it is more efficient to utilize the technology for exigency management. Mobile cloud computing provides more capabilities to create sophisticated services.

Currently available techniques do not expeditiously handle exigency notification techniques. Most of the processes predominant around the world require the victim or witness to decide the type of exigency first and then find out which departments must be contacted. For example, in case of a fire breaking out, ambulance, fire engine and police have to be called. Manually finding out the contact numbers and departments is not only ineffective, but also, at the time of exigency, it creates panic or terror and affects intellectual thinking process. This could cost life as well. A few seconds could save a lot of lives in the situation like earthquake or tsunami. Depending on application, data packet transmission in a sensor network could be periodic, event-driven or both.

Cloud computing, along with Fog computing, can play a very vital role in not only exigency alert process, but also, in the overall exigency management. Cloud computing platform provides vastly manageable and expandable virtual servers, computing resources storage resources, virtual networks, and network bandwidth, according to the requisite and affordability of customer. It also provides solution to process distributed content. Moreover, data can be accessed far and wide destitue of the trouble of keeping large storage and computing devices. Large amount of content can also be shared and cooperated easily with cloud computing.

Fog computing extends traditional cloud computing paradigm to the edge (Edge Computing). It is a Micro Data Centre (MDC) paradigm, being virtualized, able to provide computation, storage, and networking services between the end nodes and cloud environment. Fog computing is aimed at services with widely distributed deployments. Because of being localized, it provides low latency communication and more context awareness. Fog provides real-time delivery of data, especially for delay sensitive, exigency, and healthcare related services. Fog helps resource-forced nodes offload the rich tasks. It can pre-process the raw data and notify the cloud, before cloud could further adapt that data into enhanced services.

Reliable Routing Protocol (RRP) for wireless sensor networks aimed to maximize routing reliability and minimize routing overhead by discovering multiple routes from each sensor node to a sink node. Routing protocol discovers loop-free routes with each route being verified as a bidirectional communication path between a sensor node and the sink node. The route optimization tries to select disjoint routes as much as possible. Each route is uniquely identified in the network so that the node knows exactly which route to be used for sending or relaying packet. The proposed RRP provides capability to realize load balance and it can be optimized for lightweight process routing.

Although, there has been a lot of work on exigency management, but current solutions do not

address exigency notification in an effective and simple way. In this paper, we present a service architecture for exigency alert and management, through Fog, RRP and cloud computing. Our system, Exigency Alert Mobile Cloud (EAMC) tackles different kinds of exigency situations in a very simple and effective way. User only has to press a button, the application itself decides which departments have to be informed, including the location of the event happened. EAMC also informs the family members automatically, by sending messages to already stored contacts numbers.

In rest of the paper, section II is on already done work. Section III is on our proposed system. We evaluate and discuss the results of our system in section IV. Our paper concludes in section V.

II. RELATED WORK

This section discusses the already done works that are related to our area of focus, Twitter proposal by Jie Yin discuss on a system that uses natural language processing and data mining techniques to extract related situation awareness information from Twitter messages generated during various disasters and crises. System architecture is presented for leveraging social media to enhance emergency notifications. High-speed text streams retrieved from Twitter during the incidents are the data sources in this system. Again, it is dependent upon how effectively and efficiently the information is retrieved and made useful. This twitter proposal handles only the exigency related to the vehicles.

Hannes Tschofenig propose IP-based emergency service. Emergency Context Resolution using Internet Technologies (ECRIT) IETF GEOPRIV Working Group related to this sub-domain. GEOPRIV focuses on protocols and techniques required to develop a robust exigency architecture, which would be able to function on all types of IP-based networks. The caller needs to acquire location information for exigency notification, the end systems or proxies have to identify an emergency call and then mark and route it to the proper Public Safety Point (PSP). In this case, either the end host or proxy determines the location of the host. For initial location-based routing, either end host is responsible or the Session Initiation Protocol (SIP) proxy. This all process is vulnerable to location spoofing. Furthermore, a simple and quick mechanism is required for notification, which this mechanism clearly lacks.

The study discussed by Lin Dajian et al. is on formation of resources, which are required for judging exigency rescue process to have a balanced and optimized configuration of the resources. To determine the levels of the emergency response capabilities in industrial accidents, it is essential to have a balanced resource allocation, for effective rescue process. This study is mainly to overcome evaluation system from some enterprises which already exists, uncertainty factor of relevance of evaluation index system with butterfly catastrophe theory to build up four-dimensional evaluation model of the enterprise's emergency rescue capabilities. Since emergency event and its magnitude cannot be predicted, Exigency Resources play a vital role in exigency rescue process, because they are directly related to the accident's classification disposal of exigency plans. Exigency resources mainly include four types of entities: human, machine, environment, and management. Human, with some exigency training can learn exigency measures and judge potential risks involved in the accidents and the possibility of accident happening correctly and immediately. On the other hand, machine deals with the safety and facilities for handling exigency situations. Exigency channel, quantity, exigency equipment's type, performance, storage locations, and standby facilities are all that a machine has to deal with. Management is mainly mirrored in whether the whole exigency tackling process can be performed properly ,efficiently and effectively.

In another twitter proposal by Kathy Pretz discusses utility of Twitter for emergency alert. The author states that a developing situation can be assessed through the data gathered from Tweets. The prototype service in this regard uses data-mining techniques to parse high-volume Twitter streams and identify early indicators of a potential incident. The prototype service is deployed in Australia. The program continuously collects and analyzes tweets from different locations throughout Australia, using data-capture module. The authors claim that Twitter's is increasing; hence, it would be easier to report an incident. With this proposal, one issue is that when a disaster strikes, the volume of tweets can be overwhelming to be monitored and extracted. Besides, it totally depends upon those people who use Twitter and are online at that time, tweeting about that happened incident. Even in the developed countries with very high literacy rate, not everyone uses social media. Twitter's users would even be lesser then. This mechanism also has to deal with a large amount of tweets and complex algorithms. Instead, emergency management requires a more efficient and easier way, prevalent not only in Australia, but elsewhere as well, specially under-developed countries. Emergency alert process should be simple for an illiterate person as well.

III. EXIGENCY ALERT MOBILE CLOUD (EAMC) WITH FOG AND RRP

In our system, our objective mainly was to overcome the prime issue of complexity and delay in exigency notification.

Data is communicated to the Fog, which sends alert to appropriate emergency tackling departments and family members of the victim through the reliable routing protocol. Later on, data is pre-processed, filtered, and then uploaded to the cloud, which analyses it and further creates extended portfolio of services.

Among the available services, none of them is capable enough through which appropriate exigency tackling department (e.g. fire-brigade) is directly contacted by the application, upon user's single action or click of a button, instead, the user or victim has to decide which departments have to be contacted and then find out their contact numbers. Our system not only does that, at

the same time, a message is sent to the close family members of the user.

In our case, proposed EAMC maintains a list of those family members. With this, user does not need to find out which department to be communicated and search for contact numbers of family members at the time of exigency. User will only click on the type of event; rest of the things will be done by the application, in coordination with the Fog and RRP. The basic interface currently provides seven different types of emergency notification options. Figure 1 shows the basic interface.

As mentioned earlier, the data may be uploaded in the cloud, which helps related departments for better planning and betterment of the future. All concerned departments will be able to access all type of incidents' information over the cloud and analyze it. For instance, if some area experiences more accidents at nighttime due to bad light or sharp turns, then that issue can be tackled in future.

Hospitals and ambulance service providers can see which locations are more suitable to have their resource point or exigency vehicles are located, for quick response and have reachability to the place of event, keeping in view the frequency and types of events that occur in a particular area.

Figure 1.Basic interface of proposed EAMC.

In case of exigency situation, instead of thinking about whom to contact and how to contact and then inform to the family members as well one by one, the user only has to select the type of event occurred through a simple user-friendly menu. Upon doing that, the application sends message to the control center of appropriate exigency dealing department by sending a short message, including the place of that event, shown in figure 2.

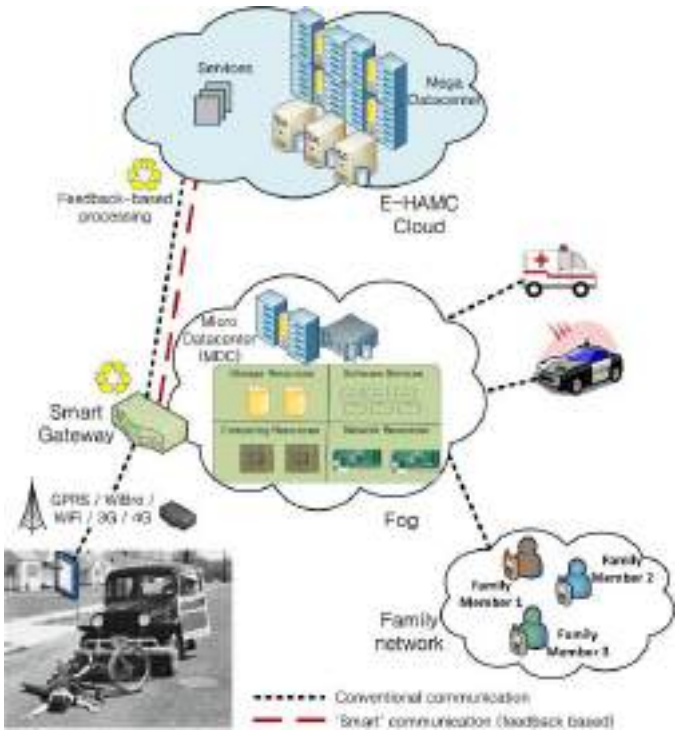


Figure 2. EAMC communication pattern.

Once the alert has been made by the Fog, data is pre-processed for further refinement and then sent to the cloud. Concerned authorities can gather the data from the cloud, when needed, to analyze which kind of exigency situations have been rising with what frequency, in any particular area and what are the reasons. This will allow preventing and avoiding such situations in future and ensure better life of a people.

A sensor node can receive multiple copies of the packet broadcasted by its neighbors. Based on the received copies of the packet, a sensor node selects two reverse routes to the sink node. The route selection criteria vary according to the different cost functions such as the hop count, the link quality, etc. In this paper, we use hop count as cost function to discover two disjoint routes as much as possible. The disjoint routes fail independently and therefore, the reliability is enhanced.

Operations in sensor networks may require bidirectional routes. In a wireless sensor network, the communication links may exhibit asymmetric properties. For some applications, it is required that the reachability of a node is verified before the routes can be used [6]. Initial route construction process uses a broadcast. The routes discovered are only valid for one direction from the sink node to sensor nodes. Therefore, routes must be verified as valid routes from sensor nodes to the sink node. The route verification also provides the sink node with downward routes to sensor nodes.

The route verification is performed by using the RV packet and the RC packet. To verify a route, a sensor node unicasts a RV packet to the sink node with DT set to 1 via the route specified by NL field in RV packet. Upon receiving a RV packet, the sink node stores route information and unicasts a RC packet back to the sensor

node along the reverse route. When the sensor node receives the RC packet, the route has been verified to be a bidirectional route.

Intermediate sensor nodes relay the RV and the RC packets. When a sensor node on the route receives a RV packet and it stores route information and forwards the RV packet to the next hop node. The stored route at intermediate node serves multiple purposes such as relaying the RC packet back to the source node of RV packet, load balance and aggregated route verification.

In aggregated route verification, a shorter route is verified as a portion of a longer route. To increase the probability of the aggregation, longer routes are preferred to be verified earlier. To perform aggregated route verification, a sensor network can estimate packet propagation time based on network diameter and wireless technology used. Before transmitting the RCD packet, the sink node can broadcast a time packet to be used by sensor nodes to estimate when route verification starts,

A Contacts should be updated according to the victim or user location

When a user has changed its location and moved to another city or country, the contacts of emergency dealing departments haveto be updated. In the new location, the application contacts the cloud and synchronizes contacts lists, along with the availability of different types of departments, dealing different sorts of disasters. By this, user never has to manually update. Users are kept always in ready state to use the application.

B. Handling of clowning appraisal in exigency notifications

Handling clowning appraisal completely may not be possible. But at least a notable effort can be made in this respect. In our system, once an exigency situations rises, the victim send picture of the event, which is then sent to the Fog automatically by the application. Since the service is going to be used with smart-phones, therefore, having camera in the equipment is not an issue. In case the victim is not in a situation to do that, any passer-by can take picture and inform the concerned departments with documentary proof. This mechanism will at least help reduce clowning appraisal, if not eliminate it.

IV. PERFORMANCE EVALUATION

The test setup was in such a way that smart phone having EAMC installed was used as end node. Our Real-time Mobile Cloud Research Center (RmCRC) private cloud XenServer was used for Fog communication. Dropbox was used as cloud storage service, where Fog uploads the data in the end using RRP through sensor nodes. Different access networks can be used in this regard. Our evaluation is based on WiFi, WiBro, Broadband, and 4G networks. Final presented results are average of all captured data analyses. The evaluation with respect to the scenarios.

In scenario 1, evaluation is on end-node to Fog communication. In scenario 2, evaluation is on end-node to cloud communication. Multimedia file data set is used to represent the situation when an audio or video file, based on the disaster, is uploaded to the cloud. In this regard, we used 20MB file for the purpose of evaluation. For different file types, different scheduling algorithms are used by the cloud. For example, shortest-job-first, first-in-first-out, etc., which have their own impact on the overall performance of data storage in the cloud. The evaluation consists of 100 instances of users, who are using EAMC service and notify for different exigency situations to the concerned departments. Results are average of all instances.

A. End-node to Fog

Table 1 shows that 20MB data communicated from end-node to Fog takes an average of 7.84 seconds. This data is then processed by the Fog (through RRP) further and after refinement, uploaded in the cloud.

Table 4.1 UPLOAD DELAY END-NODE TO FOG

DATA SIZE	20MB
UPLOAD DELAY	7.84sec

In the second form of data-set, bulk-data of different sizes was used. The result presented in table 2 shows average of 10MB bulk-data set, which contained images, location coordinates, and text. It takes 4.4 seconds to communicate 10MB bulk-data to the Fog.

Table 4.2 UPLOAD DELAY END-NODE TO FOG

DATA SIZE	10MB
UPLOAD DELAY	4.4sec

B. End-node to Cloud

This part shows that if there is no Fog involved between the end node and the cloud, the overhead is comparatively increased and delay incurred is more than the previous scenario.

Shown in table 3, uploading a 20MB video file to the cloud takes about 69.3 seconds. This is hence the average time to upload the stated size of video or multimedia data on the cloud. This shows that compared to non-Fog scenario, the incurred delay is up to 9 times.

Table 4.3 UPLOAD DELAY END-NODE TO CLOUD

DATA SIZE	20MB
UPLOAD DELAY	69.3sec

When an already uploaded content is to be relocated in the cloud or its attributes are changed, the cloud has to re-configure its URL, since every file has a unique web identity in the cloud. This relocation or change in the attributes requires synchronization. For a service being accessed by more than one node or user, collaborative environment is created, which requires more time to synchronize and update the contents. Average time to synchronize data is shown in table 4.

Table 4.4 SYNCHRONIZATION DELAY CLOUD

DATA SIZE	ALL
SYNCHRONIZATION DELAY	04sec

SYNCHRONIZATION DELAY FOR COLLABORATIVEWORK	09sec

V. CONCLUSION

Notifying incidents in an efficient way is becoming very important. In this work, we focused on the issue of quick and easy way of notifying for different sorts of emergencies or disasters. Our objective was to keep the victim from thinking too much and analyzing at the time of catastrophe. The victim or witness has to just press one button to inform about the type of event. The service automatically decides and contacts relevant departments. Fog computing and reliable routing protocol are incorporated in the model to provide resource hungry task offloading and pre-process the data. Later on, data is communicated to the cloud for more enriched services through RRP. The evaluation of the system endorses the utility of Fog in this particular scenario. Generally, with Fog, the overall delay was around six times less than the otherwise case, when data is to be directly communicated to the cloud by the end node and RRP aimed to maximize routing reliability and minimize routing overhead by discovering multiple routes from each sensor node to a sink node with network-wide broadcast flood. Routing protocol discovers loop-free routes with each route being verified as a bidirectional communication path between a sensor node and the sink node and each route is uniquely identified in the network so that the knows exactly to which route to be used for sending a packet.

REFERENCES

[1] Mohammad Aazam and Eui-Nam Huh, "E-HAMC: Leveraging Fog Computing for Emergency Alert Service" in proceedings of the 5th IEEE International Workshop on Pervasive Networks for Emergency Management,2015.

[2] Jianlin Guo, Philip Orlik, Jinyun Zhang and Koichi Ishibashi, "Reliable Routing in LargeScale Wireless Sensor Networks" Ishibashi.Koichi@ce.MitsubishiElectric.co.jp 2015.

[3] Gautam S. Thakur, Mukul Sharma, Ahmed Helmy , "SHIELD: Social sensing and Help In Emergency using mobiLe Devices", in the proceedings of IEEE Global Telecommunications Conference, Florida, USA, 6-10 December, 2010.

[4] Srikumar Venugopal, Han Li, Pradeep Ray, "Auto-scaling emergency call center's using cloud resources to handle disasters", in the proceedings of 19th IEEE Inernational Workshop on Quality of Service, San Jose, California, USA, 6-7 June, 2011.

[5] Go Hasegawa, Satoshi Kamei, and Masayuki Murata, "Emergency Communication Services Based on Overlay Networking Technologies", in the proceedings of IEEE Fourth International Conference on Networking and Services, Gosier, Guadeloupe, 16-21 March, 2008.

[6] C.E Perkins, E.M Royer and S.R Das "Ad-hoc On-demand Distance Vector (AODV) routing " http://tools.ietf.org/html/rfc3561.

[7] V.D Park and S.M Corson, "A Highly Adaptive Distributed Routing Protocol(DSR) for MANET's for IPv4" , http://tools.ietf.org/html/rfc4728.

Technique to Balance Energy Efficient Clustering with Data Transmission in Large Scale Sensor Network

Jyothi A.P, Usha Saktivel
Research Scholar,,Asst.Prof,Dept of ECE,Prof. & HOD:Dept. of Computer Science RRCE, Bangalore
jyothi_arcot@yahoo.co.in, sakthivelusha@gmail.com

ABSTRACT : Clustering has always been playing an essential role in energy efficiency over wireless sensor network. There are various studies that have been focused on resolving energy issues using various techniques where each technique have their own limitations. Hence, this paper presents a unique technique of clustering exclusively targeting large scale wireless sensor network by introducing a novel communication technique, clustering technique followed by selection of cluster head. The technique is also found to maintain a good balance between energy efficiency and data transmission in wireless sensor network. The study outcome shows proposed system offers better clustering performance with respect to most frequently adopted energy efficient protocol with respect to network delay, throughput, and energy conservation.

Keywords- Clustering, Energy Consumption, LEACH, Network Life Time, Wireless Sensor Network.

I. INTRODUCTION

A wireless sensor network is globally used for collecting information from human inaccessible geographical location. Although, the concept of wireless sensor network has shown multi-dimensional application in theory but it really doesn't even exists in commercial markets. It is interesting as there are massive research papers on wireless sensor network since last two decades, but it has never met the utility of common people very largely. A wireless sensor network consists of sensors that perform extraction of physical data from human inaccessible locations. Normally, sensors are low-powered electronic devices with low availability of resources, which are meant to capture environmental data e.g. heat, smoke, temperature, moisture, motion etc. In order to balance the tradeoff of energy consumption and network performances in a resource constraints network, it is essential to organize a multihop network into clusters. Fundamentally clustering includes two processes, one to decide the membership of participating node to a cluster family in each communication cycle and another to elect one of the nodes as a Cluster Head (CH). The selection of optimal cluster is a NP-complete problem and resembles the closeness to the "minimum dominating set problem" [1]. Energy efficiency issues are vast in literature and the reader is suggested to refer an extensive review work by the authors [2] namely "Trends and Technologies used for Mitigating Energy Efficiency Issues in Wireless Sensor Network". The prime reason behind this is a sensor node still suffers from energy problems that are the root cause of majority of other problems e.g. routing [1][2], bandwidth [3], security [4]. From the trends of existing research, one thing is very much clear that clustering is one of the most suitable factors that positively or negatively affects the network lifetime of wireless sensor network. The transmission distance of the CH is higher as compared to the other nodes, due to which they consume more energy. In order to balance this energy consumption a periodic re-cycling of CH election is adopted. A typical direct transmission from CH to base station and CH to CH as a multi-hop communication schematic is shown in the Fig.1. A direct transmission will have higher Communication overhead; it may introduce latency as well as it is not scalable for the large scale wireless sensor network as sensor could not support communication at long-haul.

The multihop communication from CH to CH having larger advantages to cope with scalability, increased life time and energy efficiency. Apart from routing based on clustering there are other secondary benefits of clustering which includes localization of the route inside the cluster so that the size of routing table within an individual node is minimized. The bandwidth utilization balancing and minimization of control message exhibit additional benefit on the overall traffic due to clustering. Irrespective of the associated advantages of clustering, the challenge of prolonging the network lifetime is crucial and still an open research problem as with multiple constraints, it is not trivial to achieve the goal by traditional approaches of clustering. The design prospect of cluster based routing protocol must include the application robustness along with secure communication, Synchronization and finally data aggregation [3].

At present, there is various energy efficient routing as well as clustering techniques introduced by various researchers till date [5]. However, all the clustering technique seriously lags inclusion of cost effective optimization (although there are various studies focusing on iterative and computationally expensive optimization theory [6] [7] in wireless sensor network). Hence, this paper discusses about a unique optimization technique called as clustering approximation that contributes to enhancing the network lifetime by providing multiple attributes for selecting cluster head. Section II discusses about the prior research work carried out in the area of energy efficient clustering in wireless sensor network followed by Section III that

briefs about the problems identification. The proposed contribution is discussed in Section IV followed by algorithm implementation in Section V. The result analysis is done in Section VI followed by summary of the paper in Section VII as conclusion.

II. RELATED WORK

This section briefs the studies being completed towards tending to energy problems in various forms of wireless sensor system. Our earlier study has investigated an existing method for solving the power issues [8]. Most as of late, utilization of Voronoi chart for vitality productivity was seen in the work of Gautam et al. [9]. The author has likewise utilized bio-inspired system for enhancing the lifetime of the network. Yu et al. [10] have exhibited uneven grouping mechanism for energy conservation on groups with unequal number of nodes. There are likewise sure studies that give one of a kind bunching system e.g. utilization of Hausdorff separation based clustering system presented by Zhu et al. [11]. Unfortunately, the results of such works are never observed to be benchmarked. Udompongsuk et al. [12] have introduced a study that performs selection of clusterhead in light of moving normal, a factual based system. Be that as it may, the study experiences versatility issue and its tradeoff with energy effectiveness. Pei et al. [13] have displayed a study where the creators have enhanced the ordinary LEACH calculation researched over intellectual radio in wireless sensor system. In any case, the system finds less convenience after some time and mission based applications in wireless sensor system. Aside from these, there are additionally different enhancement system which depends on swarm insight e.g. dolphin swarm optimization, elephant swarm optimization [14], microorganisms searching calculation [15], BAT calculation [16]. Albeit all the aforementioned studies are centered on homogenous systems, there are additionally concentrates on devoted to address vitality issues and bunching issues relating to heterogeneous system. Thinks about directed by Meenakshi [17] and Patil [18]. Albeit, all the aforementioned methods give superior rules for future examination, however every one of them is conceivably connected with issues. Brief exchange of such issues in made in next segment.

III. PROBLEM IDENTIFICATION

The past section has talked about different energy effective procedure to moderate the system lifetime and a percentage of the standard clustering techniques, which has gotten a wide acknowledgment in past. In any case, the greatest problem with every one of the systems is pretty much the same. Dominant part of the procedure considers the position of the base station at the centroid of simulation area, which prompts most extreme of routing overhead. The second issue in existing studies is selection criteria of clusterhead, which is just based on residual energy. However, a few studies have streamline energy utilizing bio-inspired procedures, yet complexities of such algorithms concerning handling time and memory utilization is missing from the latest literary works. The third issue in the latest work is the way the advancement is performed. Larger part of the cases, it is done on just single level, be that as it may, there is a reasonable probability of performing optimization utilizing different levels considering different parameters included in information aggregation of remote sensor organize separated from leftover energy. The fourth problem explored in the existing studies is consideration of uncertain knowledge about the communication behaviour of the sensors. Thus, the issue articulation of the proposed study can be characterized as - "It is a computationally difficult to build up a multi-level clustering in grouping strategy for upgrading the system lifetime in wireless sensor applications". The following segment presents around a novel method that addresses the issues.

IV. PROPOSED SYSTEM

The prime purpose of the proposed system is to develop a novel technique of clustering that can directly contribute to enhancing the network lifetime. The secondary objective of the proposed system is to ensure a well balance scheme between energy efficiency and data packet delivery. Energy efficiency is maintained by introducing a novel clustering algorithm and data delivery is maintained owing to a very unique communication model in large scale wireless sensor network. The technique introduces a concept of collection of aggregated data packets from the cluster head using multiple numbers of extra nodes called as auxiliary nodes that are assumed to be multiple sub-base stations for large scale wireless sensor network. The justification behind this is one base station will not be enough for monitoring large scale network and hence multiple base stations are used in real-time. We are bringing out novelty in the concept by introducing RF transceivers to be a part of auxiliary nodes that is mainly responsible for two task i.e. i) collecting the aggregated data from the cluster heads in proximity and ii) performing syncing with each other to identify redundant data (in order to discard them). We assume here that the RF transceivers have abundant resources and are a replacement of multihop routing in sensor network. Hence, the proposed technique offers a novel clustering technique. The technique also discusses about a simple method of cluster head selection process which is based on two parameters remnant energy and distance between nodes. The next section discusses about the research methodology that is adopted in the proposed system.

V. RESEACH METHODOLOGY

The proposed research work considers an empirical design approach for accomplishing an enhancement in network lifetime owing to novel clustering technique. The design of the proposed clustering mechanism is carried out using following modules:

- **Communication Model**: The proposed system considers a large scale simulation area of wireless sensor network with specific number of nodes distributed randomly. There are 4 types of sensors involved in the communication model e.g. i) cluster head, ii) member node, iii) base station, and iv) auxiliary node. The uniqueness in the proposed system is after the data being aggregated by the cluster head, it is forwarded to multiple auxiliary

nodes, which performs redundancy checks and then forwards the unique information to the base station. The scheme of communication model is represented in Fig.1.

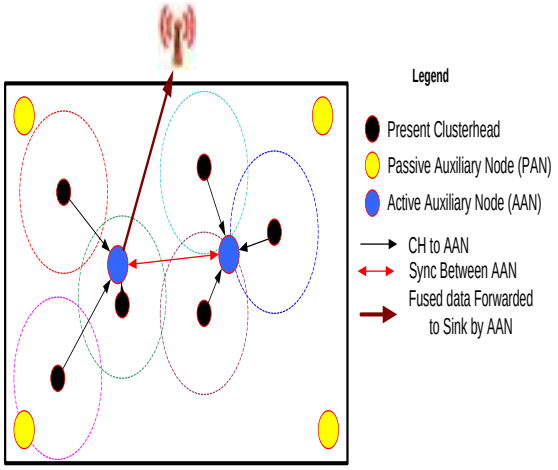


Fig.1 Proposed Communication Model

Fig.1 shows that there are two forms of auxiliary nodes i.e. i) Passive Auxiliary Node (PAN) and ii) Active Auxiliary node. Basically, an auxiliary node can be also thought as small base stations mounted in the surveillance area for coverage and connectivity extension in large scale wireless sensor network. For better energy efficiency, the system is able to identify the minimum number of such auxiliary nodes required for data aggregation from multiple spots. Hence, the technique switches of power of certain auxiliary nodes called as PAN and switches on only specific number of auxiliary nodes required called as AAN. However, with the progress of simulation, PAN and AAN toggle the switching on / off cases in accordance to coverage of present cluster head. Also the technique is also capable of performing syncing between all the AAN to ensure lower overhead in base station along with increment of data quality.

Fig.1 shows that there are two forms of auxiliary nodes i.e. i) Passive Auxiliary Node (PAN) and ii) Active Auxiliary node. Basically, an auxiliary node can be also thought as small base stations mounted in the surveillance area for coverage and connectivity extension in large scale wireless sensor network. For better energy efficiency, the system is able to identify the minimum number of such auxiliary nodes required for data aggregation from multiple spots. Hence, the technique switches of power of certain auxiliary nodes called as PAN and switches on only specific number of auxiliary nodes required called as AAN. However, with the progress of simulation, PAN and AAN toggle the switching on / off cases in accordance to coverage of present clusterhead. Also the technique is also capable of performing syncing between all the AAN to ensure lower overhead in base station along with increment of data quality.

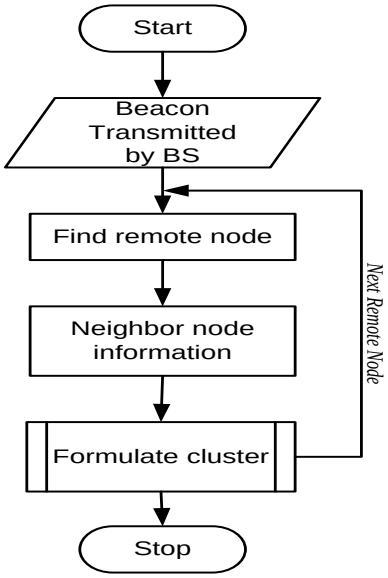


Fig.2 Proposed Clustering Techniques

VI. Algorithm Implementation

The design and development of the proposed system is carried out on the basis on 2 algorithms as discussed below:

a. Algorithm for Communication

This algorithm is responsible for performing communication in terms of data aggregation in wireless sensor network. The algorithm takes the input of *n* (total sensors), n_a (number of AAN), n_p (number of PAN), *sd* (shortest distance), *nd* (node density), bs (base station), which after processing forwards the fused data to base station. The algorithm initially performs random distributed of nodes and make uniform distribution of n_a and n_p. Using the second algorithm for the clustering, the algorithm aggregates the data from cluster head and forwards it to n_a. The selection mechanism of n_a is based on shortest distance between the cluster head and uniformly positioned auxiliary nodes. However, with a progress of simulation, it is quite possible that existing cluster head will be depleted of battery power and in that case possibly the older n_a will be not be enough to cover the range of new cluster head. Hence, we calculate the node density in order to check the positions of majority of nodes. The algorithm also re-calculates the distance between the newly elected cluster head with all n_p and n_a (which are reachable). The n_a which are in reachable distance are retained and other are replaced by new n_a which was previously n_p. After the complete collection of the data is accomplished, the communicating na will perform syncing with each other, which is basically an operation to find the message with similar timestamp. Such message are considered to be duplicated message and hence discarded. Finally, the data packets are fused together by anyone of the na which is forwarded to base station.

Table 1 Algorithm for Communication

Algorithm for Communication

Input: *n* (total sensors), n_a (number of AAN), n_p (number of PAN), *sd* (shortest distance), *nd* (node density), bs (base station)

Output: fused data forwarded to base station

Start

1. init rand(n), n_a, n_p
2. Apply Algorithm-2 for CH selection.
3. n_a sd(CH, auxiliary node)
4. If battery(CH)==0
5. Calculate n_d and sd[(newCH),(n_a, n_p)]
6. Select n_p new(n_a)
7. $n_{ai} \leftrightarrow$ fuseddata(bs) $\leftrightarrow n_{aj}$
8. n_a filter(fuseddata) b_s.

End

The advantage of this algorithm is that it reduces the complexity operations over cluster head that results in maximizing the lifetime of the cluster head to 50%.

b. Algorithm for Clustering

The algorithm selects the input of *bs* (base station), *d* (distance), E_c (Energy Coefficient), and E_{rem} (remnant energy) and after processing gives the output of energy efficient cluster head. This algorithm is also used in previous algorithm of communication. This technique initiates with a base station bs to broadcast the route discovery information so that it can evaluate the position of other nodes from itself along with other status of explored nodes. The base station than uses Euclidean's distance between themselves and other nodes and then it selects the node with maximum Euclidean's distance, which means that this node is located in more remote area within the simulation area. We call it as last node that starts exploring the neighborhood nodes. It does so by using distance-based formulations i.e. if the transmission distance of the next node is found to be within the present node, the next node is considered to be neighbor nodes. The process is repeated for next node with lesser distance from the last node. Hence, a cluster is formulated where all the parent nodes *i* is transmitted with the data from the children node *j*. After the cluster is formulated the next emphasis is on the selection process of cluster head. A new variable called as energy coefficient E_c is computed to be remnant energy of each node divided by distance between the sensor and base station. The system than consider a node with the maximum value of E_c to be cluster head. This algorithm, when used with communication algorithm, ensures highest network lifetime along with delivery of quality data during the data aggregation process in wireless sensor network.

Table 2 Algorithm for Clustering

Algorithm for Clustering

Input: *bs* (base station), *d* (distance), E_c (Energy Coefficient), E_{rem} (remnant energy)

Output: Selection of clusterhead

Start

1. bs broadcast(beacon)
2. $d \rightarrow \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \; \forall (x_1\, y_1) \in^n (x_2\, y_2) \in bs$
3. For i=1:n
4. Last node max(d)
5. Last node find_neighbor_node
6. Next node [max(d)-a]
7. Next node find_neighbor_node
8. Repeat Line-5 until all nodes are covered.
9. Calculate energy coefficient

   $E_c = E_{rem} / d$
10. Clusterhead max(E_c)
11. End

End

Hence, it can be also seen that there are multiple points of n_a where the data aggregation takes place from cluster head to na and while multiple n_a sync with each other to delivery redundant free data to sink. Hence, data aggregation time may quite increase. So, we first compute the total time required to capture the data by all n_a for the purpose of enhancing the network lifetime. Hence, we develop a condition which ensures that data aggregation time from n_a should be minimized. As the existing cluster head is free from performing redundancy check or multi-hop operation, hence, it network lifetime is maximized. The next section discusses about the results accomplished from the study.

VII. RESULT DISCUSSION

The outcome of the proposed study was compared with all the standard energy efficient protocols e.g. LEACH [19], PEGASIS [20], ERP [21], HCR [22], and HEED [23]. These are the frequently used energy efficient clustering technique for wireless sensor networks. The accomplished outcomes of the study are discussed as follows:

A. Analysis of Energy Conservation

The amount of energy conserve is calculated by finding total amount of average residual energy retained by the nodes in 1000 simulation rounds. The outcome shown in Fig.3 highlights that proposed system offers better retention of energy as compared to existing system. HCR algorithm is designed using genetic algorithm for optimizing the energy efficient routes. Usage of genetic algorithm over longer iteration has extra computation of

fitness function and extracting the elite outcomes. Hence it includes more energy consumption just like LEACH. The performance of LEACH and ERP is degraded owing to overheads as clusters once selected as allowed to work until death. Moreover, as base station is always located at center, it incurs more loss of power in case of redundant data packets. The improvement of LEACH could be seen in PEGASIS; however, the chains don't supports extensive clustering mechanism with redundancy check on data packet.

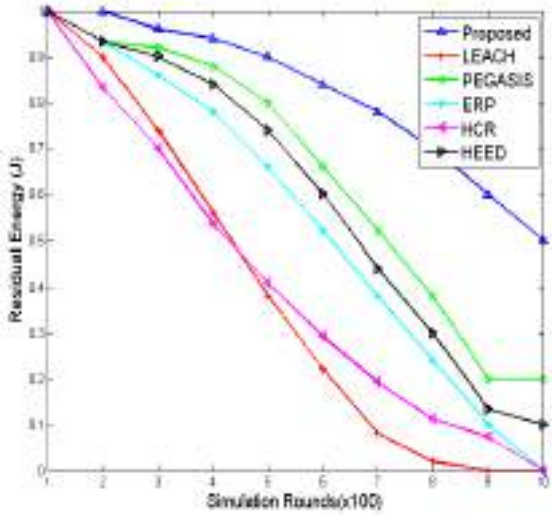


Fig.3 Energy Consumption Analysis

A. *Analysis of Throughput*

Throughput was estimate by calculating the size of data packet being processed (fused) and transmitted over certain specific period of time. Fig.4 shows that LEACH, HEED, and ERP doesn't have much better throughput performance. Owning to usage of conventional topology, the task of redundancy check (data fusion) has to be carried out by each cluster head. Hence, in case of dynamic wireless sensor network, the system for existing technique fails to perform intercommunication among the clusters in order to check for unique packets. As such algorithms don't supports multihop network, hence, a good amount of transmits energy was expended in order to maintain good coverage and communication performance in LEACH, HEED, and ERP. A closer look will also show that PEGASIS and HCR has nearly similar performance for throughput. The prime reason behind this is HCR uses round robin scheme and PEGASIS uses chain-based scheme for performing clustering. Being a part of hierarchical routing, PEGASIS and HCR scheme ensures better energy efficiency but it considerably takes time to construct the path which is a time consuming process especially for HCR which uses iterative evolutionary technique. Hence, proposed system excel better as it uses AAN and PAN, which acts as subsidiary base station for large scale network, for perform the complex computational work by synching with each other, which is just a replica of multihop communication system.

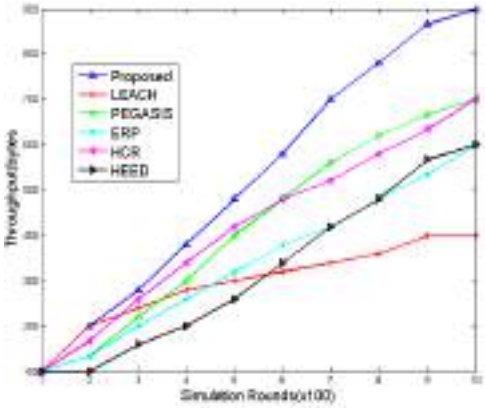


Fig. 4 Throughput Analysis

B. *Analysis of Network Delay*

Network delay is computed by the time difference between the packets being relayed to packets being received. It can be easily computed by the observing the timestamps of packets in transmitting state and receiving state. Fig.5 highlights the analysis of the network delay, where it can be seen that PEGASIS has been witnessed with more network delay. The prime point of issue in the chain structure, which is good for energy efficiency but is not mean to meet the communication objectives of time critical applications in wireless sensor network. The next better performance of delay was seen in trend of ERP and HEED due to usage of optimization principle. Better delay performance was also observed for HCR that uses genetic algorithm. However, it was not found energy efficient.

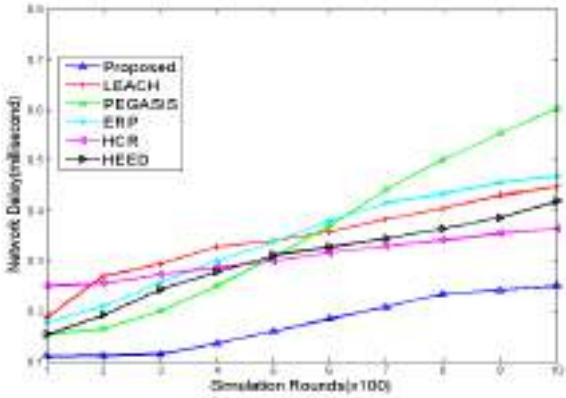


Fig.5 Network Delay Analysis

Therefore, the outcome clearly shows that proposed system is capable of ensuring minimization of energy drainage, maximization of data delivery, and with highly reduced network delay.

VIII. CONCLUSION

Energy is constantly the uncommon resource in any applications on vast scale remote sensor framework and grouping is the most as often as possible utilized procedure to ensure the life range of the sensor hubs. Till 10 years prior, there have been tremendous abstract works towards vitality change in bunching. Out of such rules research responsibility, a couple of systems have gotten a basic affirmation and are comprehensively packed in association with vitality powerful gathering instrument.

Regardless, such existing models encounters i) nonappearances of more broad degree of progression, ii) utilization of an over the top measure of complex count that impacts correspondence execution and life range of sensors, and iii) openness of lesser number of energy capable grouping measures. The best pitfalls of existing clustering methodologies is its absurd suspicions i.e. i) determination of sink on the premise of leftover vitality, ii) constraining the base station in the point of convergence of recreation territory (i.e. position reliance), iii) nonattendance of supportability of multihop correspondence in vitality effective methods. Hence, this paper addresses all such basic issues by showing a novel batching method that performs potential streamlining. Various criteria have been point by point for decision of aggregator center point, which gives better edge to vitality insurance. Using first demand radio model, the energy assessment shows that proposed grouping change technique ensure perfect life range of a sensor hubs when contrasted with existing framework

REFERENCES

[1] M. Ghaffari, "Near-optimal distributed approximation of minimum-weight connected dominating set", In Automata, Languages, and Programming, Springer Berlin Heidelberg, pp. 483-494, 2014

[2] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey",Computer networks, Vol. 54(15), pp.2787-2805,2010

[3] H. Jiang, L. Chen, J. Wu, S. Chen, and H. Leung, "A reliable and high-bandwidth multihop wireless sensor network for mine tunnel monitoring", Sensors Journal, IEEE, Vol. 9(11), pp.1511-1517,2009

[4] G. Padmavathi, and Shanmugapriya, "A survey of attacks, security mechanisms and challenges in wireless sensor networks", arXiv preprint arXiv:0909.0576, 2009

[5] S.K. Singh, M.P. Singh, and D.K. Singh, "A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks",International Journal of Advanced Networking and Application (IJANA), Vol. 2(02), pp.570-580,2010

[6] K. Selvarajah, and V. Kadirkamanathan, "Energy efficient sink node placement in sensor networks using particle swarm optimization", In Ant Colony Optimization and Swarm Intelligence, Springer Berlin Heidelberg,pp. 510-511, 2006.

[7] H. Safa, and F. Yassine, "Localization in large scale wireless sensor networks. In Telecommunications (ICT),2012 19th International Conference, pp. 1-6, 2012.

[8] A.P. Jyothi and U. Sakthivel. "Trends and Technologies Used for Mitigating Energy Effi-ciency Issues in Wireless Sensor Network", International Journal of Computer Applica-tions, vol.111, No.3, pp.32-40, February 2015.

[9] N. Gautam, S. Sofat, and R. Vig, "An Ant Voronoi Based Clustering Approach for Wire-less Sensor Networks", Springer Journals, Social Informatics and Telecommunications, 2014

[10] J. Yu, Y. QI, G. Wang, "An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks", Springer Journal of Control Theory Application, 2011

[11] X. Zhu, L. Shen, and T-S Peter Yum, "Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks", IEEE

[12] Transactions On Vehicular Technology, vol. 58, no. 2, February 2009

[13] K. Udompongsuk, C. S. In, C. Phaudphut,, "MAP: An Optimized Energy-Efficient Cluster Header Selection Technique for Wireless Sensor Networks", Springer Journal, Advances in Computer Science and Its Applications, Lecture Notes in Electrical Engineering, 2014

[14] E. Pei, H. Han, Z. Sun, B. Shen and T. Zhang, "LEAUCH: low-energy adaptive uneven clustering hierarchy for cognitive radio sensor network", Springer- EURASIP Journal on-Wireless Communications and Networking, 2015

[15] M.A. Bharathi, B.P. Vijayakumar, D.H. Manjaiah, Cluster Based Data Aggregation in WSN Using Swarm Optimization Technique, International Journal of Engineering and In-novative Technology, Vol.2, Iss.12, June 2013

[16] M.A. Bharathi, M. Mallikarjuna, B.P. Vijaya Kumar, "Bio-Inspired Approach for Energy Utilization in Wireless Sensor Networks", Elsevier-International Conference On Model-ling Optimization And Computing, Vol.38, pp.3864-3868, 2012

[17] K. Seelam, M. Sailaja and T. Madhu, "An Improved BAT-Optimized Cluster-Based Routing for Wireless Sensor Networks", Springer Journal, Intelligent Computing and Ap-plications, Advances in Intelligent Systems and Computing, 2015

[18] D. Meenakshi and S. Kumar, "Energy Efficient Hierarchical Clustering Routing Protocol for Wireless Sensor Networks", Springer Journal, Social Informatics and Telecommunications Engineering, pp.409-420, 2012

[19] P. R. Patil and U. P. Kulkarni, "Energy-Efficient Cluster-Based Aggregation Protocol for Heterogeneous Wireless Sensor Networks", Springer-Journal, Intelligent Computing, Net-working, and Informatics Advances in Intelligent Systems and Computing, 2014

[20] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocols for Wireless Microsensor Networks", Proceedings of the 33rd Hawaaian International Conference on Systems Science (HICSS), 2000

[21] C.S. Raghavendra "PEGASIS: Power-Efficient Gathering in Sensor Information Systems Stephanie Lindsey",Work, Vol. 310, pp. 336-1686,2001.

[22] E. Ahvar, S. Ahvar, G.M. Lee, and N. Crespi, "An energy-aware routing protocol for query-based applications in wireless sensor networks, The Scientific World Journal, 2014

[23] S. Hussain, and A.W. Matin, "Energy efficient hierarchical cluster-based routing for wireless sensor networks", Jodrey School of Computer Science Acadia University Wolfville, Nova Scotia, Canada, Technical Report, pp.1-33,2005.

Efficient Energy Conservation Technique Using Sink Node Mobility for Big Data Gathering in Densely Distributed WSN

Karthik K J, Anitha K
PG Scholar, Rajarajeswari College of Engineering, Bengaluru -74
karthikkj1992@gmail.com, anithakrishna14@gmail.com

ABSTRACT- Big data has considered being promising application in the field of information and communication technology (ICT). WSNs are used to gather big data in communication technology. In most of the case the data from single sensor is not enough informative to user. To gather significant data, it is necessary to gather information from distributed sensors. Because of limited power in WSNs big data gathering is challenging task. To increase network lifetime in WSNs the paths for data transfer are chosen in such a way that the total energy consumed along the path is reduced. To support improved data aggregation and high scalability, sensor nodes are often grouped into non-overlapping, disjoint subsets called clusters. Clusters create hierarchical WSNs which incorporate efficient utilization of limited resources of sensor nodes and hence extends lifetime of network. In this paper we proposed an efficient way for data gathering using clustering based technique. In the proposed system we used k-medoids clustering technique for the optimal number of clusters and created sink node trajectory, used as efficient data collector. And also using novel approach for data encryption and decryption for providing security.

Keywords - Big data, Data aggregation, clusters, k-medoids, sink node trajectory.

I. INTRODUCTION

Wireless sensor networks (WSNs) have come up with an effective solution for many of the applications like military, healthcare, home, etc. recent advances in wireless communications and electronics have enabled the development of inexpensive, low power and multi-utility sensor system. A WSN consists of several number of sensor nodes that are distributed densely within a network area. The position of sensor nodes are need not be predetermined or engineered. This allows random distribution in inaccessible terrains or disaster relief operations. On the other hand, this also means that WSN network protocols and algorithms possess self-organizing capabilities. Another unique feature of sensor networks is the cooperation of sensor nodes. Sensor nodes are fitted with an onboard processor. Instead of sending the unprocessed data to the nodes are responsible for further action, they use their on-going abilities to carry out simple computations and move only the required and partially processed data. The described features ensure a broad range of applications for WSNs. Some of the applications are health, military and home. In military, for example, the rapid deployment, self-organization, and fault tolerance of sensor networks make them very promising sensing techniques for military operations, control, interactions, performing, intelligence, surveillance and targeting systems. Implementation of above mention and other WSNs applications require wireless ad hoc networking techniques. Protocols used in wireless ad hoc networks were not well suited the unique features and application requirements of sensor networks. The major limitations of wireless sensor networks their limited power storage capacities and prone to failures. Hence energy efficiency is the major issue in sensor network.

As information technology grows rapidly, volume of the data also increasing simultaneously. Accessing the retrieving big data is crucial for the user in many applications. Big data is a buzzword, or catch-phrase, used to describe a massive amount of both structured and unstructured data that is so huge that it's tough to process using database and software techniques. In most of these scenarios the data is too large or it moves rapidly or it exceeds present processing capacity. Big data has the potential to help companies improve operations and make faster, more intelligent decisions. Collecting large amount data from sensor nodes is the major concern in the field of ICT. Individual sensor nodes may not provide accurate information. Therefore collecting data from multiple sensor nodes is very essential.

In order to gather these data, the WSNs are constructed in such a way the sensors relay their data to the "sink". However, in case of widely and densely distributed WSNs there are two problems in gathering the data sensed by billions of sensors. First, the network is divided to some sub-networks groups because of the limited wireless communication range. Second, the wireless transmission consumes the energy of sensors. Even though the size of data generated by an individual WSN is not significant, each sensor requires a lot of energy to process the data formed by surrounding sensors. Especially in dense WSNs, the life time of sensors will be very less because each sensor node compute a lot of data generated by enormous number of surrounding sensors. In order to overcome these problems, we need an energy-efficient method to gather huge volume of data from a large number of sensors in the densely distributed WSNs.

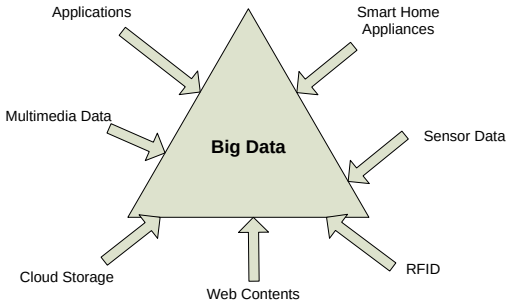


Fig 1. An Overview of Big Data Gathering

To achieve energy-efficient data gathered in densely distributed WSNs, there have been many existing methods. Clustering is one of the most widely used techniques to make WSNs energy efficient. Low energy adaptive clustering hierarchical (LEACH) clustering algorithm is introduced to maximize the life time of wireless sensor network. In [] used expectation maximization clustering technique.

In this paper we proposed an improved method to gathering large data in densely distributed sensor network using k-medoids clustering algorithm. K-medoids algorithm is used as clustering of WSNs. K-medoids is not only effective in clustering also robust in selecting cluster heads in large WSNs. Later we introduced an effective sink node routing for collecting data from each cluster heads..

II. RELATED WORK

[5]The promise of data-driven decision-making is now being recognized broadly, and there is growing enthusiasm for the notion of "Big Data". Heterogeneity, scale, timeliness, complexity, and privacy problems with Big Data impede progress at all phases of the pipeline that can create value from data. The problems start right away during data acquisition, when the data tsunami requires us to make decisions, currently in an ad hoc manner, about what data to keep and what to discard, and how to store what we keep reliably with the right metadata. Much data today is not natively in structured format; for example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display, but not for semantic content and search: transforming such content into a structured format for later analysis is a major challenge. The value of data explodes when it can be linked with other data, thus data integration is a major creator of value. Since most data is directly generated in digital format today, we have the opportunity and the challenge both to influence the creation to facilitate later linkage and to automatically link previously created data. Data analysis, organization, retrieval, and modeling are other foundational challenges. Data analysis is a clear bottleneck in many applications, both due to lack of scalability of the underlying algorithms and due to the complexity of the data that needs to be analyzed.

[4]In this paper, we formulate a flow control optimization problem for wireless sensor networks with lifetime constraint and link interference in an asynchronous setting. Our formulations based on the network utility maximization frame work, in which a general utility function is used to characterize the network performance such as throughput. To solve the problem, we propose a fully asynchronous distributed algorithm based ondual decomposition, and theoretically prove its convergence. The proposed algorithm can achieve the maximum utility. Extensive simulations are conducted to demonstrate the efficiency of ouralgorithm and validate the analytical results.

[2]This paper presents and analyzes three-tier architecture for collecting sensor data in sparse sensor networks. Our approach exploits the presence of mobile entities (called MULEs) present in the environment. When in close range, MULEs pick up data from the sensors, buffer it, and deliver it to wired access points. This can lead to substantial power savings at the sensors as they only have to transmit over a short-range. This paper focuses on a simple analytical model for understanding performance as system parameters are scaled. Our model assumes a two-dimensional random walk for mobility and incorporates key system variables such as number of MULEs, sensors and access points. The performance metrics observed are the data success rate (the fraction of generated data that reaches the access points),latency and the required buffer capacities on the sensors and the MULEs. The modeling and simulation results can be used for further analysis and provide certain guidelines for deployment of such systems.

[3] Clustering is a standard approach for achieving efficient and scalable performance in wireless sensor networks. Traditionally, clustering algorithms aim at generating a number of disjoint clusters that satisfy some criteria. In this paper, we formulate a novel clustering problem that aims at generating overlapping multi-hop clusters. Overlapping clusters are useful in many sensor network applications, including inter-cluster routing, node localization, and time synchronization protocols. We also propose a randomized, distributed multi-hop clustering algorithm (KOCA) for solving the overlapping clustering problem. KOCA aims at generating connected overlapping clusters that cover the entire sensor network with a specific average overlapping degree. Through analysis and simulation experiments we show how to select the different values of the parameters to achieve the clustering process objectives. Moreover, the results show that KOCA produces approximately equal-sized clusters, which allows distributing the load evenly over different clusters. In addition, KOCA is scalable; the clustering formation terminates in a constant time regardless of the network size.

[8]A WSN is a specialized wireless network made up of a large number of sensors and at least one base station. The foremost difference between the WSN and the traditional

wireless networks is that sensors are extremely sensitive to energy consumption. Energy saving is the crucial issue in designing the wireless sensor networks. Since the radio transmission and reception consumes a lot of energy, one of the important issues in wireless sensor network is the inherent limited battery power within network sensor nodes. In order to maximize the lifetime of sensor nodes, it is preferable to distribute the energy dissipated throughout the wireless sensor network. The data gathering schemes should be power efficient. In our proposed work we are changing the idea related to the data gathering and transmission protocol Chiron. The main goal of our research is reduce of energy consumption and improve the lifetime of network as chain leader belonging to the certain covering angle will only transmits the gathered data to the another chain leader of the same covering angle and then we send the data of the another covering angle in sequential manner . So the data is transferred to some angle based chain leader rather than to the nearest chain leader. By this method of data gathering we found that energy consumption is reduced and lifetime is improved significantly.

III. PROBLEM STATEMENT

There are three Problems in gathering the data sensed by millions of sensors.

1. Network is divided to some sub-networks because of limited wireless communication range.
2. The wireless transmission consumes the energy of the sensors during data gathering.
3 There should be only one sink node for each sub-network, which reduces delay in data-gathering.

.

IV. METHODOLOGY

1. Network initialization.
2. Cluster formation.
3. Calculating cluster centroids.
4. Identifying sink node and evaluating trajectory.
5. Efficient data gathering using sink node.
6. Result analysis.

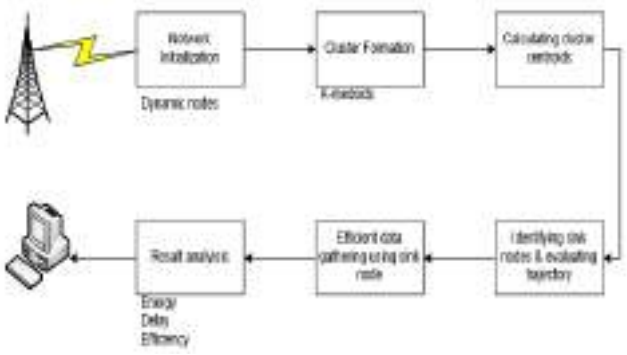


FIG 2: PROPOSED ARCHITECTURE.

V. ALGORITHM

Input: the network graph$G(V, E)$; k, the number of clusters to be formed; λ, the fading factor of physical channel

Conditions: a set of clusters involving all nodes

1: Pick k nodes randomly as initial cluster heads, and use M to denote the set of those k cluster heads.

2: Establish an empty set C_i for head h_i of$M$, and initialize

3: Establish a set

4: while any node u of V/M do

5: u chooses h_i of M as its head such that

Where d_{u,h_i} is the Euclidean distance from u to h_i.

6:$C_i \leftarrow C_i \cup u$

7: end while

8: while any C_i do

9: Select a node u from C_i as new head such that

$$\arg_u \min\left\{\sum_{v=1,v\neq u}^{|C_i|} d_{u,v}^{\lambda} | \forall u \in C_i\right\}$$

10: $N \leftarrow N \cup u$

11: end while

12: if $N \neq M$ then

13: $M \leftarrow N$

14: $N \leftarrow \emptyset$

15: go to step 4

16: end if

VI. PERFORMANCE EVALUATION

The following graphs are compared to the existing method which are based on Expectation-Maximization technique (EM-METHOD) to the K-medoids technique. The graphs are showed better performance in energy, delay and data delivery efficiency.

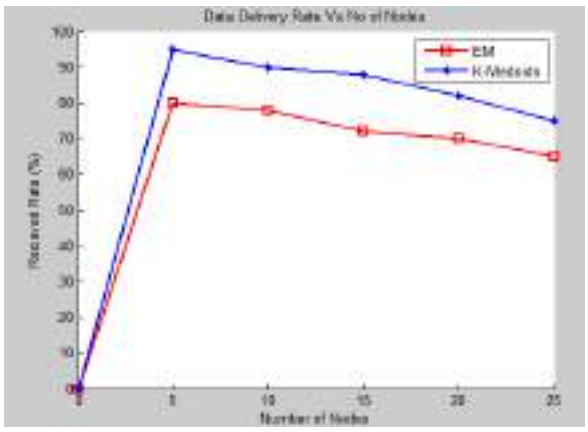


Fig 3: Data Delivery Performance

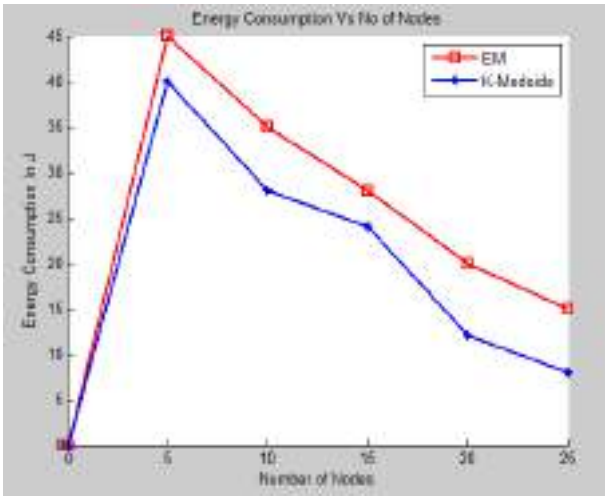


Fig 4: Energy Level Performance

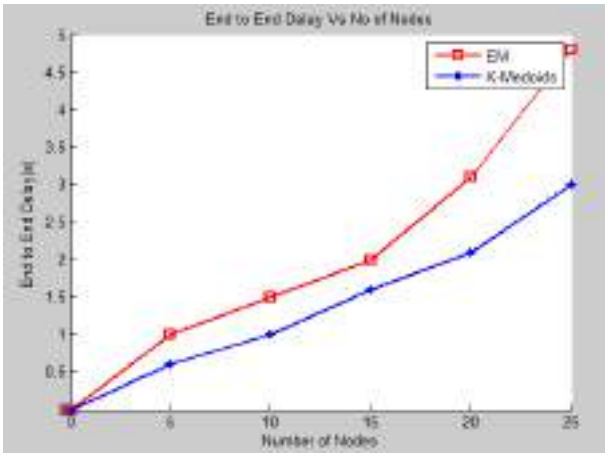


Fig 5: End to End Delay Performance

VII. CONCLUSION

The system gathers big-data energy efficiently using this proposed scheme for large-scale wireless sensor networks. This system suggests that energy efficient big data gathering in such networks is, indeed, necessary. We use optimal routing algorithm for communicating between sink node and cluster head. Once the cluster head gathers information from each node within a cluster, then that information should be passed to sink node in single-hop path. If the communication range is far than normal then multi-hop path is taken for gathering data. By using optimal routing mechanism we can gather big-data energy efficiently and also by using novel approaches it is possible to secure the data effectively from eavesdroppers..

ACKNOWLEDGEMENTS

We would like to thank the management and faculty of RRCE, CSE Dept.

REFERENCES

[1]. C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and powercontrol in a multicell wireless data network," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 1883–1892, 2001

[2] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: modeling and analysis of a three-tier architecture for sparse sensor networks," Ad Hoc Networks, vol. 1, no. 2-3, pp. 215 – 233, 2003

[3]. M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 20, no. 12, pp. 1844–1856, Dec. 2009.

[4] C. Jiming, X. Weiqiang, H. Shibo, S. Youxian, P. Thulasiraman, and S. Xuemin, "Utility-based asynchronous flow control algorithm for wireless sensor networks," IEEE Journal on Selected Areas in Communications, vol. 28, no. 7, pp. 1116–1126, Sep. 2010.

[5]. A. Divyakant, B. Philip, and et al., "Challenges and opportunities with Big Data," 2012, a community white paper developed by leading researchers across the United States.

[6] E. Begoli and J. Horey, "Design Principles for Effective Knowledge Discovery from Big Data", Software Architecture (WICSA) and European Conference on Software Architecture (ECSA) Joint Working IEEE/IFIP Conference on, Helsinki, August 2012

[7] Intel IT Center, "Peer Research: Big Data Analytics", Intel's IT Manager Survey on How Organizations Are Using Big Data, August 2012.

[8] Parminder Kaur,Mamta Katiyar, "Energy-Efficient Chain-Based Data Gathering in WSN," International Journal of Science, Engineering and Technology Research (IJSETR) Volume 1, Issue 6, December 2012.

IMPLEMENTAION OF PRIVELIGE DATA IN CLOUD COMPUTING BY DOUBLE ENCRYPTION CONCEPT

Govindaswamy H R, Bhanu K N
Department of MCA, Rajarajeswari College of Engineering, Bangalore, India
govindaswamy.hr@gmail.com,bhanu.kn@gmail.com

ABSTRACT-This paper implements secure data in cloud computing by double encryption method using RSA. Double Encryption for secure outsourcing data in the cloud. This method solves key escrow problem and Data reveal problem by RSA algorithm of asymmetric key approach. In existing mCL-PKE method, there is Certificate-less Encryption and also single encryption and it is half disrupted by the "CLOUD" and remaining half is decrypted by the user but in my method there will be certified for the User and two layer encryption. The cloud will going to decrypt the outer layer encryption and user will going to fully decrypt inner layer encryption only for doing this we can secure the data / information well be highly secured.

Index Terms— DEA, Decryption center, RSA algorithm, cloud computing, Asymmetric key, Two Layer Encryption

I. INTRODUCTION

In this newly environment of double encryption concept. certification of the users provide high security to the data and asymmetric key approach (RSA) is very convenient in key distribution. The implementation method is that RSA can also be used for performing digital signature. In Existing system Two Layer Encryption and it is extended from the previous method of mCL-PKE. mCL-PKE works on certificate-less encryption and user are not certifie by any authorize entity, but in my method, there will be certified for user, certification of the user also provides security of the information in the cloud, due to this only permitted user can use the data as mentioned in [1],[2]. The Double Encription Approach (DEA) this two layer encryption approach addresses the drawbacks of the mCL-PKE . In DEA approach user will have to get register to the participants to get the secret key for decryption of the encrypted documents. The basic method is, owner encripts the documents and send these encrypted documents to the cloud, now cloud decript the outer-layer of the encripted content and send these document to the register users, now user fully decripts the encrypted contents ,it shows an inner layer of the encription of the secret keys.

II. SYSTEM ARCHITECTURE

Fig1.system architecture

In this paper the proposed method is divided into three main parts they are:

1) Owner

2) Cloud

3) User

Cloud is further divided into three subparts Encrypted Storage(ES),DecriptionCenter(DC)KeyGenerationCenter(KG)ThecloudhasthreesubpartsEncriptedContentStorage,K eyGenerationCenter(KGC),andSecurityMedeationServer(SEM).Encripted Content Storage is going stores the encrypted documents KeyGenerationCenter,is going to generate the KGC key for encryption and Security Mediation Server partially is going to decrypts the encrypted information.Cryptography is the art and science of achieving security bye encrypting/encoding. And also it used for securing the data to non readable longuage,the process of encoding plain text messages into secret message is called as encryption, for encrypt the data there are many techniques are there. Encryption is one of technique to protect the data from spiteful and not registered users, encryption of the documents can be more than one layer.

III. RSA ALGORITHM

In double encryption technique the RSA algorithm is one of the best algorithem.On this algorithm we are going to use two important keys that private key and public key,public key is used for encrypting the information in the form of non readable language and private key is used for decrypting the encrypted document/information. the In the real-world public-key cryptosystems RSA algorithm is one of the best algorithm and is widely used to secure data transmission. In such a cryptosystem, the user can easily view the encryption key because it is public but decryption key is kept as top-secret. In RSA, this asymmetry is based on the real-world difficulty of distributing the product of two large prime numbers. Basically RSA algorithm works slow. and because of this

commonly used to directly encrypt user data as mentioned in [5] RSA pass the encrypted shared keys for symmetric key cryptography, which can start doing encryption-decryption of information very fastly.observance is the important thing in RSA algorithm impotent thing is to find the three large positive integers that is d,e and n such that with modular exponentiation for all m and even knowing m and n even m can be extremely difficult to find d.

IV. DOUBLE ENCRYPTION

The implementation of double encryption method is to secure the TEXT Data Items of the Data-Owner by Double Encryption in Cloud Computing. The basic method is Double Encription of the documents means there is two layer encryption of the data orinformation. I extend the previous mCL-PKE method but in my system there is certification of the users. The simple method is owner will encrypt the contents two times using thegenerated key and store the document to theEncripted Storage when user request any document thedecription center fetche the requested document and decrypts the outer layer of encryption and gives to the user,now user fully decrypts the document.

In this section I propose the basic mCL-PKE method then my improved method, the basic public key encriptioniscertificate-less method, in which users certificationisnotnecessarywhichreducesthemanagementcost.Butthis method compromises to the malicious users, any malicious user can access the data for malicious use. The shortcomings of this is addressed by the improved method in my paper, in which user must have to register to the owner then only user can access the information. So this ideology enhances the security of the data. The basic mCL-PKE method is going to do single encryption and half decrypted by the cloud and remaining information will be decrypted by the user, this method is proposed to reduce the decryption time of the user, but partially decryption of the data reduce the security of the content, but in this method there is double encryption of the data, there is two layer of the encryption, in which the cloud will going to encrypt thr outer layer decrypted by the cloud and inner layer encryption is decrypted by the user, hence security is high in my improved method. The overall result comes that securityis very high in my system as compare to previous mCL-PKE method.

.

V. IMPLEMENTATION OF DOUBLE ENCRYPTION PROCESS

Fig 5 shows the user registration form user as to register with administrator .the owner will going to give the permission to the user to access the file. In our paper the user request the owner for registration purpose and the same user request the decryption centre for decrypting the key which is provided by key distribution centre. Owner plays an important role in our project, he request key from key distribution center for a user and encrypts that key and sends it to decryption centre. The owner also stored the doubly encrypted file. After registration the user choose file and encrypt that file use the key provided by owner and stored the encrypted file in cloud.

Fig2.user registration form

Fig3.owner home page

Ofter the user is registered the owner will going to check who are all the clints are registerd and the owner will going to send the keys to that user for doing their work.

```
mysql> desc data_owner;
+---------------+-------------+------+-----+---------+-------+
| Field         | Type        | Null | Key | Default | Extra |
+---------------+-------------+------+-----+---------+-------+
| owner_name    | varchar(22) | NO   |     | NULL    |       |
| owner_address | varchar(50) | NO   |     | NULL    |       |
| sex           | varchar(10) | NO   |     | NULL    |       |
| contact       | varchar(12) | YES  |     | NULL    |       |
| email         | varchar(30) | NO   |     | NULL    |       |
| department    | varchar(30) | NO   |     | NULL    |       |
| password      | varchar(15) | NO   |     | NULL    |       |
+---------------+-------------+------+-----+---------+-------+
7 rows in set (0.01 sec)
```

Fig4.description of data owner table

The above figure shows the description of data owner table this table contains owner name,adress contact, email and password .it cintains the complete details of owner .ofter owner registration completed the owner details will be stored in this table.

```
mysql> desc owner;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key | Default | Extra |
+----------+-------------+------+-----+---------+-------+
| owner_id | varchar(15) | NO   |     | NULL    |       |
| password | varchar(15) | NO   |     | NULL    |       |
+----------+-------------+------+-----+---------+-------+
2 rows in set (0.04 sec)
```

Fig5.description of owner table

The above figure contain the owner_id and Password. The owner_id and Password will be stored in owner table.

Fig6.verifyication page

In this form verifying the user information whether the user has enterd correct email and mainly encrypted key and also offer verifying the details it generate the decrypt key. contains owner name, dress contact, email and password .it contains the complete details of owner .ofter owner registration completed the owner details will be stored in this table.

VI. Conclusion

The implementation of the double encryption process is going to provide certification for the user to highly secure the data or information. In this process we are using an asymmetrical Key approach that is Ron Rivest, Adi Shamir, and Leonard Adleman (RSA) algorithm for key distribution. The future enhancement of this double encryption process and RSA can use for performing digital signature and it has also helped full for providing high security in future. In this method I'm using basic Mediated certificate less public key encryption (mCL-PKE) key for Key distribution and also I'm using two important key that is public and private key .public is for encrypting the document / information in the form of non-readable language means it will going provide security for that document or information. The shortcomings of this method are addressed by the improved method in my system, in which user must have to register to the owner then only user can able to access the information. So this ideology enhances the security of the data.

VII. Acknowledgement

The authors of this paper are thankful to the Management, RRGI,Bengaluru and Principal, RRCE, Bangalore for their continousencouragement and support throughtout the work.

References

[1].Mr. Bhavesh Rahulkar received the BE (ComputerTechnology) From RTM Nagpur University, Nagpur (M.H.) in 2008Mr.
[2]. Praveen Shende, Asst. Prof.,CSE Dept. C.S.I.T.Durg, India, received B.E. (Computer Sc.) in year 2009
[3]. Mohamed Nabeel, Elisa Bertino, Seung-Hyun Seo, Xiaoyu DingMembers of IEEE "An Efficient Certificate-less Encryption for Secure DataSharing in Public Clouds" June 2013.
[4]. Zhiguo Wan, Jun"e Liu and Robert H. Deng. Senior Member, IEEE"HASBE: A Hierarchical Attribute-Based Solution for Flexible and ScalableAccess Control in Cloud Computing" April 2012.
[5]. https://en.wikipedia.org/wiki/RSA_(cryptosystem).

Govindaswamy H R is currently as PG student of Rajarajeswari College of Engineering pursuing final year of Master of Computer Application.

Bhanu K N is currently working as Associate Professor in the Department of Computer Applications. Her research and professional career spans for about 12 years of Teaching and Industrial Experience. She obtained her MCA., from IGNOU, New Delhi, M.Phil., in Computer Science from VMU, Salem and presently Pursuing Ph.D., in Computer Science from Rayalaseema University, AP. Her expertise is primarily in the domains of Wireless Sensor Networks, sensor clouds, Adhoc Networks and Sensor Big data analytics.

Secured Load Re-Deployment of File Chunks in Distributed File System

Roopadevi D S, Mrs. Nandini G B.E,M.Tech(Ph.D)
PG Student, Assistant Professor ,Department of Computer Science,
RajaRajeswari College of Engineering, Bangalore-74
roopadevids@gmail.com, Email: nanduamma@gmail.com

ABSTRACT-In cloud computing environment, the Distributed File Systems are the most important building blocks. The DFS is used to store nodes and they perform many functions of computing applications. Files can be created, deleted and appended dynamically. It results in load imbalance in a distributed file system, that is, file chunks are not distributed as uniformly as possible among the nodes. The distributed file system in production system powerfully depend on the central node for chunk re allocation. This dependency is clearly inadequate in a large scale; failure-prone environment because the load balancer in the center is put under significant workload that is linearly scaled with the system size and because of that, it becomes the performance bottleneck and the single point failure. A fully distributed file load balancing algorithm is proposed to handle the load imbalance problem. Load is transferred from heavily loaded node to physically closed lightly loaded node. Encryption is done for the raw file before dividing it into chunks and later it will be decrypted.

Keywords - Distributed File System, Name node, Data Node, Cloud Computing, Map Reduce.

1. INTRODUCTION

Distributed file system is mainly based on the client-server architecture, where the client obtains the data stored on the server as if it is locally present on their own system. The distributed file system is mainly developed for extensive data storage and data access.

The main characteristic is that the file contents can be stored in various nodes which shares and transfer the data. There are various types of distributed file system, one among them is the HDFS (Hadoop Distributed File System).

HDFS is very similar to the Google File System. HDFS is built using java language. HDFS is designed to store and process large data sets. It is designed in such a way that it can run on the computer hardware which is affordable and easy to obtain. HDFS takes care of storing files of huge volume of data. HDFS component is used by the administrator. The main features of the HDFS are:

- It is fault tolerant and can be developed on the low cost commodity hardware.
- It provides high throughput admission to application data.
- Quick and automatic recovery from the faults.

HDFS applications are based on the notion of "write one read many". HDFS architecture is like master-slave architecture. The HDFS consists of two main components namely, A single name node and the multiple instances of data node. The name node is also called as master server or admin node. Remaining all other nodes are called data nodes. These Data nodes are also called as slave node or worker node. The name node and data nodes are software's designed to run on the low cost hardware machines.

A file is split or divided into a number of small pieces called chunks. These chunks of files are assigned to the different data nodes. The data nodes create or delete the file chunks in accordance with the instruction of the central node or main server.

The functionalities of the name node are:

- It manages the metadata information of the file system.
- It controls the access to files by client.
- It allows mapping of data blocks to data nodes.
- It performs operation like opening, closing and renaming of files.
- It provides the list of HDFS files that belong to each data block, the current location of the block and the state of the file.

The functionalities of data node are:

- Information contained in the files is stored in the data nodes, that is, the file contents are stored in the data nodes.
- Data nodes serve the read and write requests from the clients.
- It performs creation, deletion and replication of the blocks based on the instructions from the name node.

In, HDFS, the file is divided or split into number of small pieces called chunks. These chunks are assigned to different data nodes. The load possessed by the nodes is proportional to the number of file chunks the data node holds. A HDFS cluster is said to be in a balanced state if there are no overloaded or under loaded based on the percentage of the DFS space used by the data nodes.

A data node is said to be overloaded data node if the DFS space used by the data node is greater than the predefined threshold. A data node is said to be under loaded if the DFS space used by the data node is lesser than the predefined threshold. The threshold can be

configured by the user that is, the threshold can be changed by the user.

The imbalanced state is created in the HDFS as and when the new data nodes joins the system. Whenever the file is created and also when the file chunks possessed by the data nodes exceed the threshold capacity or when the file chunks possessed by the data nodes are below the threshold capacity. The distributed file system depends on a central node for chunk redistribution. This dependency puts the central node under the considerable workload and becomes the performance bottleneck that causes the entire process to slow down or stop and single point of failure.

The objective is to allocate the chunks of the files as equally as possible among the storage nodes such that no nodes manage an excessive number of chunks and also movement cost is reduced. To achieve the objective, a load redistribution algorithm is used to solve the load imbalance problem among all the chunk servers that is the data nodes in the HDFS. Data nodes performs the load rebalancing task spontaneously without name node due to this the performance of the utilization of the system is improved.

2. LITERATURE SURVEY

Paper 1: Hsueh-Yi Chung, Che-Wei Chanz Hung-Chang and Hsiao Yu-Chang Cho [1] proposed the Load Rebalancing Problem in Distributed File Systems where the Distributed File Systems are principal and the crucial building blocks for cloud computing applications. In such a file systems, nodes concurrently assist storage and computing functions. A file is divided into a number of small pieces called as file chunks and are assigned in evident nodes so that venture can be performed in analogous over the nodes. In a cloud computing environment, failure is normal and nodes maybe enhanced, restored and added in the system. Files can also be dynamically deleted, created and added. The outcome of this is load imbalance state, that is, the file chunks are not distributed as evenly as possible among the nodes. Despite of distributed load balancing algorithm exist in the literature to deal with the load imbalance problem. Upcoming distributed file system in production as systems completely depend on the central node for chunk re assignment. This dependence is clearly insufficient in a large scale failure and vulnerable environment because the central load balancer is placed under considerable workload, that is, it grows linearly with the system size and thus becomes the staging bottleneck and the single point of failure.

A file is divided into a number of small pieces called as file chunks and are assigned in evident nodes so that venture can be performed in analogous over the nodes. In a cloud computing environment, failure is normal and nodes maybe enhanced, restored and added in the system. Files can also be dynamically deleted, created and added. The outcome of this is load imbalance state, that is, the file chunks are not distributed as evenly as possible among the nodes. Despite of distributed load balancing algorithm exist in the literature to deal with the load imbalance problem. Upcoming distributed file system in production as systems completely depend on the central node for chunk re assignment. This dependence is clearly insufficient in a large scale failure and vulnerable environment because the central load balancer is placed under considerable workload, that is, it grows linearly with the system size and thus becomes the staging bottleneck and the single point of failure.

Paper 2: Prasanna Ganesan, Mayank Bawa and Hector Garcia [2] proposed Online Balancing of Range-Partitioned Data with Applications to Peer-to-Peer Systems where the range partitioning is used to solve the problem of horizontally partitioning a dynamic relation over a large number of disks of nodes. Such a kind of segregation is usually advisable in large scale parallel databases, as well as in peer-to-peer systems. As and when the tuples are deleted and inserted, the segregation may need to be modified and data migrated in order to achieve storage balance across the participant disks or nodes. Efficient asymptotically optimal algorithms that confirm storage balance at all times are used, even against a characterized deletion and insertion of tuples. The above algorithm is consolidated with distributed routing structure to architect a peer-to-peer system that supports efficient range queries, while concurrently assuring storing balance.

Paper 3: H.-C. Hsiao, H.Liao and S.-S. Chen [3] proposed Load Balance with imperfect Information in Structured Peer-to-Peer System with the conviction of virtual server, peers involved in a heterogeneous, structured peer-to-peer network may host different numbers of virtual servers, peers can balance their loads corresponding to their loads corresponding to their capacities. The already present decentralized load balance algorithms designed for the heterogeneous, structured peer-to-peer networks either construct additional networks to exploit global information or demand peer-to-peer nodes to be organized in a hierarchical fashion. Without depending on any backup networks and independent of the geometry of the peer-to-peer substrates. Based on the partial knowledge of the system, A novel load balancing algorithm that is distinct estimates the probability of distribution of the capacities of peers and loads of virtual servers, causing in imperfect system condition, peers can determine their anticipated loads and reassign their loads in parallel. The comparison is done through notable simulations; it is differentiated with prior load balancing algorithm.

Paper 4: Wenqiu Zeng, Ying Li and Jian Wu [4] proposed load rebalancing in Large-Scale distributed File System with the advancement of data on the internet has shown the massive growth. Some researchers have paid their attention to find an effective way to keep and govern these data. A load rebalancing algorithm to solve the load balancing problem all compute nodes in distributed file

system is used. It also guarantees that one block of file and its two replicates are stored in three different chunk servers simultaneously at the same time. This algorithm fulfills the load balancing while guarantying the reliability of the system.

Paper 5: Qingqing Zhong [5] proposed a load-balancing approach for DHT-based P2P Networks the nitpicking issue in the efficient operation of peer-to-peer network is load balancing. There are various number of suggestions exist for load balancing for structured peer-to-peer network. The adaptive load balancing for structured peer-to-peer system is used. The technique aims to balance the request and routing load of the peer under unfairly request of workloads. These procedures outstandingly improve the distribution of the load and provide inevitably better scalability. With more workloads the false positives rate is reasonably low through experiment results.

3. PROPOSED SYSTEM

In this paper, in the proposed system a load rebalancing algorithm is used to tackle the load imbalance problem. This algorithm is made to run in the rebalancing server which makes the rebalancing decision. It is invoked with the command as and when the imbalance state is created in the system. This rebalancing algorithm makes each and every data node present in the HDFS to communicate with each other and estimate the load of the neighboring data nodes. Based on the estimated load the data nodes are classified as overloaded data node and under loaded data node. These data nodes are sorted based on the number of file chunks each data nodes possesses. The data nodes are sorted from least lightly loaded node to heavily loaded node. The least lightly loaded node migrates the chunk of file it possess to the next lightly node to transfer the file chunks which is lesser than or equal to the threshold. This process repeats unless and until all the data nodes in the system becomes normal loaded. The Raw data file is encrypted before dividing the file into chunks and it will be decrypted again later.

3.1 Benefits of Proposed System

- The chunks are assigned as equally as possible among the data nodes.
- The movement cost is reduced.
- It solves the performance bottleneck problem and a single point of failure as it no longer depends on the central nodes for chunk reallocation.
- The data nodes balance their loads spontaneously by eliminating the dependency on the central node.
- The performance of the system is improved.
- The resource availability is improved and resource utilization also increases.

In the Proposed system four different module are implemented. They are as follows,

- Chunk Distribution
- Identifying Nodes
- Load Rebalancing Algorithm
- Security

3.2 Load Rebalancing Algorithm Module

In the rebalancing algorithm, every chink storage node initially estimates whether or not it is lightly loaded or heavily loaded node. A node is lightweight if the amount of chunks it hosts is smaller than the predefined threshold. Every node contacts variety of other nodes within the system and builds a vector denoted by V. this vector consists of entries and every entry contains the IDs of the nodes. This light weighted node selects one of the heavy weighted node for the reallocation of the chunks that heavy weighted node possesses.

Chunk creation

Begin
Select a file to split
For the selected file
Split the file into chunks
For end
Store the file chunks into the servers
End

Chunk Servers Module

Begin
Interact with each server to gather information regarding light and heavy load, including locations of chunks.
If under loaded then send request to migrate the chunks
Repeat till no progress
End

3.3 Stepwise Implementation of Rebalancing Algorithm:

1. To Estimate and find the lightly loaded node in the set of sample data nodes

Step 1: Get the live data nodes report from name node.
Step 2: Calculate the average utilization of DFS space used.
Step 3: Sort the data nodes according to the DFS space used in ascending order.
Step 4: Choose the least lightly loaded node as source node i in sample data nodes.
Step 5: Choose the heavily loaded node as destination node j in sample data nodes.

2. To migrate file chunks across the data nodes

Step 6: i moves the file chunks it possesses to i+1.
Step 7: i requests j to migrate the chunks it possesses below or equal to the threshold.
Step 8: i becomes j's successor that is i←j+1.
Step 9: j removes the file chunks assigned to i.
Step 10: Estimate the load of the sample data nodes.
Step 11: If imbalanced goto step-3 repeat the process unless and until all the sample nodes are balanced.

Step 12:update the chunk location information to name node.

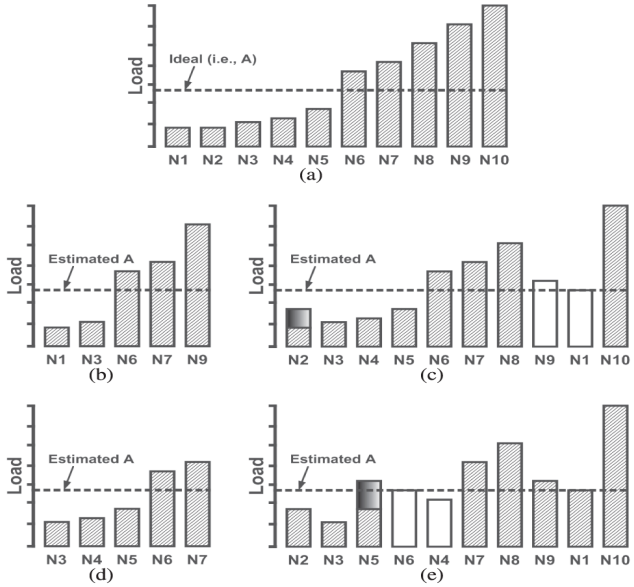


Fig 1. An example illustrating our algorithm where (a) the initial loads of chunk servers N1,N2,N3,…..N10, (b) N1 samples the load of N1,N3,N6,N7 andN9 in order to perform the load rebalancing algorithm, (c) N1 leaves and sheds its loads to its successor N2 and then rejoins as N9'ssuccessor

We have to make sure that the data that has to be stored in the cloud has to be secured. In order to provide the security, the raw data is encrypted. In this proposed system, an algorithm is used to encrypt the data. Later this encrypted file is divided into chunks and stored in various nodes.

Fig. 2. Before: Imbalanced loads on Data Nodes

Fig. 3. After- Balanced loads on DataNodes

CONCLUSION

This paper focuses the load rebalancing algorithm that solves the load imbalance problem in distributed file system. It also tries to balance the loads of nodes and minimizes the migration cost by taking the information of the node locality. It solves the performance bottleneck problem by self modifying the loads of the nodes without depending on the central node to balance the load. It also improves the outcome of the system meanwhile improves the resource availability and resource utilization.

The algorithm plays vital role in the performance of the system. The algorithm attempts to balance the loads of the file chunks as much as feasible, since the data nodes spontaneously does the rebalancing task without the interference of the name node. It not only accomplishes the load balancing but it also makes sure that the system is reliable.

ACKNOWLEDGEMENT

I am very much thankful to **Dr. Balakrishna**, Principal, RajaRajeshwari college of Engineering, Bangalore for giving this prestigious opportunity of being a member of this institution.

I am very proud and thankful to **Dr. Usha S,** HOD, Department of Computer Science of Engineering for her valuable advice and ideas at various stages of this work.

I render my whole hearted thanks to my project guide **Mrs. Nandini G** Assistant Professor, Department of Computer Science and Engineering, who has been guiding staff to achieve the goal.

REFERENCES

[1]. Hsueh-Yi Chung Che-Wei Chanz Hung-Chang Hsiao Yu-Chang Chao, "*The Load Rebalancing Problem in Distributes File System*", 2013.

[2]. Prasanna Ganesan, Mayank Bawa and Hector Garcia-Molina, " *Online Balancing of Range-Partitioned Data with Applications to Peer-to-Peer Systems*", 2012.

[3]. H.-C. Hsiao,H. Liao, s.-s Chen , "*Load Balance with Imperfect Information in Structured Peer-to-Peer Systems*", April. 2011.

[4]. Wenqiu Zeng, Ying Li, Jian Wu, "*Load rebalancing in Large-Scale Distributed File System*" , 2009.

[5]. Qingqing Zhong, " *A Load-Balancing Approach for DHT-Based P2P Networks*", 2009.

[6]. G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall and W. Vogels, *"Dynamo: Amazon's Highly Available Key-Value Store*", Oct.2007.

[7]. Wang Yue, Cai Wangdong, Duan Qi, "*An Adaptive Dynamic Load Balancing Algorithm*", 2006.

[8]. Yingwu Zhu and Yiming Hu*," Efficient Proximity-Aware Load Balancing for DHT-Based P2P System*" , April 2005.

[9]. D. KArger and M. Ruhl, " *Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems*," June 2004.

[10].J.dean and S.Ghemawat," *MapReduce: Simplified Data Processing on Large Clusters*", Proc.Sixth Symp.Operating System Design and Implementation(OSDI '04),pp.137-150,Dec 2004.

[11].P.Jamuna and R. Anand Kumar "*Optimized Cloud Computing Technique to Simplify Load Balancing*" International Jornal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013.

Enhancing Computer Inspection Using Document Clustering for Analysis

Mr.Chandrasekhar M S1, Mr.Subhash B N2, Mr.Satish V3, Mr.Bharath J4
1, 2, 3 UG Student, Dept of CSE, RRCE, 4Asst.Prof, Dept of CSE, RRCE, Bangalore-74
chandrashekarms214@gmail.com, subhashbn71@gmail.com, RRCE,sathishv138@gmail.com,
bharath5911@gmail.com

ABSTRACT-In document analysis, Computers having huge amount of data files really creates disorder to analyze it, most of the data consist in those files will be unstructured whose analysis will be difficult. Therefore, we present an approach that reduces the effort of analysis by clustering the document. Clustering is a division of data into groups of similar objects. The clustering techniques used in our approach are k-means and incremental mining, both this algorithm facilitate in discovering new and useful information from the documents under analysis. Finally, we show our results in graphs for better summarization and visual presentation purpose.

Keywords- Document clustering, text mining, Analysis

I. INTRODUCTION

The most common style of unattended learning is by clustering and this is often the major distinction between clustering and classification. No super-vision implies that there's no human professional who has allotted documents to categories. The goal of a document cluster theme is to reduce intra-cluster distances between documents, whereas increasing inter-cluster distances (using AN applicable distance measure between documents). A distance measure (or, dually, similarity measure) therefore lies at the center of document clustering. The large form of documents makes it nearly impossible to form a general algorithmic rule which might work best just in case of all types of data

Many times, it takes a lot of time to scan all the computer information and look for the required file. Thus, a professional usually scans the computers manually and tries to gather needed data. But, it may take lots of efforts and time a protracted time. Thus, so as to beat this drawback, the idea of document cluster is terribly helpful. The clustering algorithms are often helpful wherever no information concerning the information in connected document square measure best-known a priori [2],[3]. Thus, clustering helps a lot to partition information into cluster of connected documents. There are numerous cluster approaches with well-known algorithms like k-means, k-medoid, single link, complete link, etc. [1]. In our proposed system, we've used k-means and incremental mining algorithm. K-means algorithms works on comparatively validity index to estimate the cluster numbers automatically and incremental mining algorithm uses cosine similarity measurement. We've also targeted on preprocessing steps like removal of stop words and to stem the words which may facilitate to form the data which will be organized. Thus, smart preprocessing techniques will facilitate to scrub and create that data to be effectively utilized in cluster method. Our proposed methodology forms the most clusters wherever connected documents found. We tend to confer our results with the help of graph for higher summarization and visual presentation purpose.

II. LITERATURE SURVEY

There are studies regarding use of clustering algorithms in the field of text analysis of documents. Most of the studies describe the use of algorithms for clustering data e.g., k-means, k-medoid, Fuzzy C-means(FCM), single link, complete link Expectation-Maximization (EM) for unsupervised learning of Gaussian Mixture Models, and Self- Organizing Maps (SOM). SOM [2] is generally similar to K-means but are usually less efficient. SOM based algorithms were used for clustering files and making the decision-making process performed by the examiners more efficient and accurate. The files were clustered by taking into consideration their creation dates/times and their extensions.

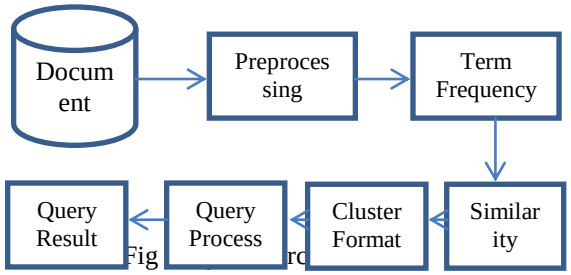
Jian Ma, Wei Xu [8], proposes an Ontology based Text Mining (OTMM) method to cluster research proposals in a research funding agency. SOM is applied to cluster the research proposals on the basis of similarity. After the grouping, reviewers are assigned with the proposals. Hence this approach reduces the time of grouping the research proposals. Qiujun [12], proposes an approach for extraction of reports content using similarity measure based on edit distance to separate the news content from clattering data. This paper describes regarding the correct extraction of reports content from web pages. The algorithms used with this methodology are less advanced with high accuracy and efficiency rate. S. C. Punitha, M. Punithavalli[11], studied two approaches for text clustering and compared them. First method relies on pattern recognition with semantic driven ways for clustering text documents. Second method is ontology based mostly text clustering approach. Each algorithm are analyzed in terms of efficiency and speed of clustering, however the performance of ontology based approach was higher in terms clustering quality, but lack in speed.

III. PROPOSED SYSTEM

In our proposed system we have decided to choose two main clustering algorithm i.e. k-means and Incremental mining algorithm. These algorithms run with different combinations of their parameters, resulting in different algorithmic instantiations. The aim is to dynamically create the cluster, carry out the analysis and to measure the performance of used algorithms. The methodology is

to accept the data from user including large sets of text documents, apply the pre-processing on the text file then clustering it using clustering algorithms. The clustering is done with reference of sample document which will be match with all other documents and clusters are formed by analyzing the difference between documents and the centroids used in clustering algorithms.

IV. SYSTEM ARCHITECTURE



Fig

Before providing documents to clustering algorithm, we tend to perform some preprocessing steps. Specially, Stop-words such as prepositions, pronouns, articles are removed. Then, we tend to adopt a conventional applied mathematics approach for text mining, within which documents are described in a vector space model. During this model, every document is described by a vector containing the frequencies of occurrences of words that are outlined as delimited alphabetic strings, whose range of characters is between four and twenty five. We also used a dimensionality reduction technique called Term Variance (TV) that may increase both the effectiveness and potency of clustering algorithms. TV selects variety of attributes (in our case a hundred words) that have the best variances over the documents. So as to calculate distances between documents, two measures are used, namely: cosine-based distance and Levenshtein-based distance. In order to estimate the amount of clusters, a wide used approach consists of obtaining a group of data partitions with totally different numbers of clusters and so choosing that specific partition that gives the most effective result according to a particular quality criterion (e.g., a relative validity index). Such a collection of partitions might result directly from a hierarchical clustering dendrogram or, or else, from multiple runs of a partition algorithm (e.g., K-means) ranging from totally different numbers and initial positions of the cluster prototypes. Query process is simply like search engine in the web, the results are the based upon the page ranking of the documents, similarly in our system results (i.e. documents) are show based on term frequency.

V. IMPLEMENTATION DETAILS

A. Preprocessing Steps:

Preprocessing of text documents is important to clean data and provide algorithms solely the desired data. It takes input as a text document and output a collection of tokens (which may be single terms or n-grams).

The preprocessing techniques used in our system are described below:

1. **Tokenization:** It takes text as input and outputs the number of tokens.
2. **Removal of Stop Words:** we tend to maintain a stop word table having all potential stop words. We scan our documents to seek out such stop words and take away it.
3. **Stemming**: Once stop word removal, we performed stemming of words. We maintained indexed stems. For first index position we kept the original stem, and then we tend to scan the document to form the stems. For example: bail / bailed / bailing. So, if we found any word like bailed or bailing then we replace these with bail
4. **Weighted matrix construction:** It involves the development of weighted matrix primarily based upon the frequency of occurrence of words.

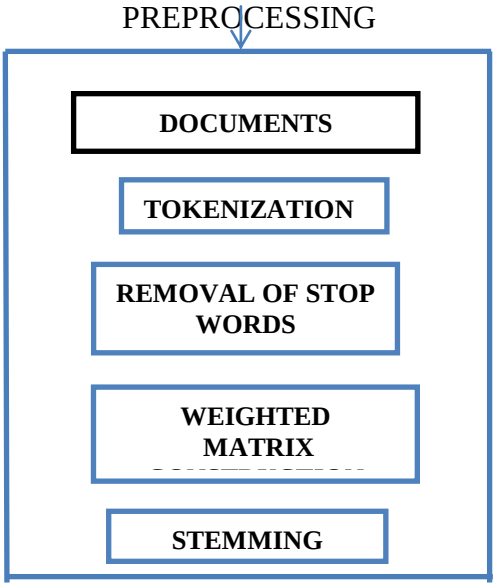


Fig 2, Preprocessing of Unstructured Documents

B. Clustering Algorithms:

K-means and Incremental mining are the famous Algorithms in the machine learning and data mining fields, and therefore they have been used in our study.

K-means Algorithm:

K-means starts with selection of K randomly chosen objects as initial clusters centers, named as seeds. The cluster centers are moved around in space in order to Minimize the RSS. These two steps are repeated iteratively until a stopping criterion is met.

- Reassignment of objects is done to the cluster with the closest centroids.
- Each centroid is recomputed based on the current members of its cluster.

The termination conditions as stopping criterion are:

- The numbers of iterations are equal to a pre-decided value for number of iterations to be completed.
- The centroids μi are not toggling between iterations.

- Termination of algorithm when the RSS value falls below a pre-established threshold.

Algorithm for K-Means

1. **Procedure** KMEANS(X,K)
2. {s1, s2, · · · , sk} Select Random Seeds(K,X)
3. **for** i ←1,K **do**
4. μ (Ci) ← si
5. **end for**
6. **repeat**
7. mink~xn−~μ(Ck)k Ck = Ck [{~xn}
8. **for all** Ck **do**
9. μ(Ck) = **1**
10. **end for**
11. **until** stopping criterion is met
12. **end procedure**

Cosine Similarity Measure: Measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them.
The cosine of 0° is 1 and for any other angles it is less than 1. It is a judgment of orientation and not of magnitude:

- Hence, the two vectors with the same orientation have a cosine similarity of 1.
- The two vectors at 90° and two vectors diametrically have a similarity of 0.
- Opposed have a similarity of - 1independent of their magnitude.

Cosine similarity is used in positive space particularly, where the outcome is neatly bounded in [0, 1].

VI. ADVANTAGE

• *Finding Similar Documents*: This feature is commonly used once the user has noticed one "good" document during a search result and desires more-like-this. The fascinating property here is that clustering is able to find documents that are conceptually alike in contrast to search-based approaches that are solely able to discover whether or not the documents share several of similar words.
• *Organizing Large Document Collections*: Document retrieval focuses on finding documents relevant to a specific query, however it fails to resolve the problem of creating sense of an oversized variety of unsorted documents. The challenge here is to arrange these documents in taxonomy just like the one humans would produce given enough time and use it as a browsing interface to the original collection of documents.
• *Search Optimization*: Clustering helps plenty in rising the standard and potency of search engines because the user query may be initial compared to the clusters rather than comparing it directly to the documents and therefore the search results may also be organized easily.

VII. CONCLUSION

The paper presents the analysis of documenting clustering techniques like partitioned clustering and hierarchical clustering. K-means presents the partitioned clustering and Hierarchical clustering makes an attempt to form a hierarchical decomposition of the given document collection therefore achieving a hierarchical structure. Similarity measures can be used to outline the performance of clusters. This approach may be very helpful for organizing vast unstructured information into structured information and enhance the method of document examination.

VIII. REFERENCES

[1] L. Filipe da Cruz Nassif, "Document clustering for forensic Analysis" IEEE Transaction on Information forensics and Security, 2015.
[2] B.S Everitt, s. Landau, and M. Leese, "cluster Analysis," London U.K 2005.
[3] A. K. Jain and R.C Dubes, "Algorithm for clustering data".
[4] C. M Bishop, Pattern Recognition and Machine Learning. New York 2006.
[5] E. R. Hruschka, R. J. G. B. Campello "an approach for extraction of news content using similarity"
[6] N. L. Beebe and J. G. Clark "Digital forensic text string searching, Elsevier 2012.
[7] Hadjidj, M. Debbabi, H. Lounis, "Towards an integrated e-mail forensic analysis frame-work" Elsevier, vol. 5, 2009.
[8] Jian Ma, Wei Xu, Yonghong Sun, Efraim Turban, Shouyang Wang, and Ou Liu, ― An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection,‖ IEEE Trans. Syst., Man, Cybernetics., A, Syst., Humans, vol. 42, no. 3, pp. 784–790, May. 2012.
[9]. S. Decherchi, S. Tacconi, J. Redi "text clustering for digital forensic analysis" Intell. Security Inf. Syst. 2009
[10]. "Document Clustering for computer Inspection" IJETR JAN 2015.
[11] Punitha, S. C., and M. Punithavalli, "Performance Evaluation of Semantic based and Ontology Based Text Document analysis", Elsevier 2012.
[12] Qiujun, L. 2010, "Extraction of News Content for Text Mining Based on Edit Distance", Journal of Computational Information Systems, 2010, pp.3761-3777.

ANALYSIS AND IMPLEMENTATION OF MULTICAST VIDEO USING OPTICAL NETWORK BY STREAMING PROCESS

[1]Nalina kumara, [2]Bhargav Rama Gowd, [3]Hitesh Mehta
[1,2]Department of ECE TOCE, Bangalore. Karnataka state, India.
[3]Managing Director for Eagle Photonics Pvt Ltd Karnataka state, India

ABSTRACT- The performance and analysis of Multicast groups using optical cable of 1KM distance network based on RTP(Real time Transfer Protocol) for the communication on real time connection oriented network Multimedia communication in internet need a large Bandwidth, Viewers or Peoples are more interested in watching live programs. Thus here we are using the RTP, During the transmission of the data from server to client systems redundant bit are introduced leading to buffering in order to overcome this streaming is adopted. Transmitting Multicast data from one point of server to the interested receiver or groups receiver systems, The main aim of this project is to improve the life time of network, Qos, communication system, Increases Bandwidth and Essential to avoid packet loss, delay than the unicast communication.

Keywords-Addresses of IP, Multicast, Optical cable, Streaming Process, Videos.

I. INTRODUCTION

The popular application in internet is a Multimedia communication. It needs to larger bandwidth a unicast is a point to point communication and acts a single direction or single host to overcome this here we taken Multicast group communication, Multicast environment consist of data is send server to multiple group of receiver is called multipoint communication or Multicast group, it gives packets are delivered high quality and The Qos(quality of service) is dependents on computer network. A Multicast address starts from 224.0.0.0 to the 239.255.255.255. the membership of hosts. Local area network of a the translation address between multicast traffic is sent over a switched Ethernet network. In this we used IGMP (Internet Group Management protocol) when TTL =1 then the IGMP is a joining the Multicast groups to hosts, and it is transmits the Membership Report Messages to the router system. For given multicast packets. IGMP messages are never forwarded by routers provides a path to receiver groups of Multicast computer systems.

A Multicast address is chosen for the receives in a multicast group. The membership of hosts can join and leaves at any time, A multicast group can be active for a long time network system are converters the routers systems, In multicast data communication the router system are used the IGMP. Multicast is delivery of High quality video, Bandwidth Efficiency more than unicast communication, During the transmission of the data from server to client systems redundant bit are introduced leading to buffering in order to overcome this streaming is adopted, the network is performed the live programs like cricket, audio and video conferences and also live radio programs ect.

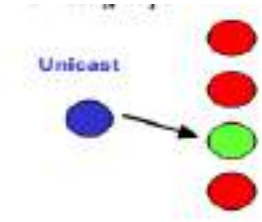


Fig1: unicast communication.

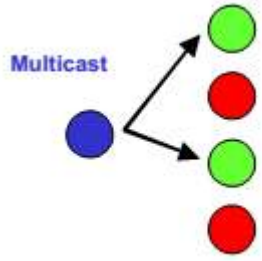


Fig 2: Multicast communication.

Fig 1 and fig 2 Comparison of unicast and multicast communication from server to client system a shown in above figs. Peoples are more want to watch live programs like audio, video conferences, cricket program, live News ect.. here Multicast results is more efficient than unicast.

II.SCOPE AND MOTIVATION

To maintain an unwanted flow of data also Quality of service(Qos), it is to avoid packet delay in the multicast data stream. Applying Qos and giving Multicast data packets priority over other packets. The multicast stream usually have some buffering built in so that losses are smoothed out than the unicast communication. The data is transfer from one point to point communication or single directions. A service. where data is delivery from a sender to a multiple receiver groups is called

multicast communication. Main aim of this project is it eliminates redundant bits and gives the High quality of results, The more important gives the long distances communication by using Optical cable.

III. IGMP MULTICASTCOMMUNICATION

Internet group management protocol (IGMP) supports to multicast communication. It does not send messages to router system. IGMP is informs to router for receive multicast host or multicast packets from the given Multicast address of host.

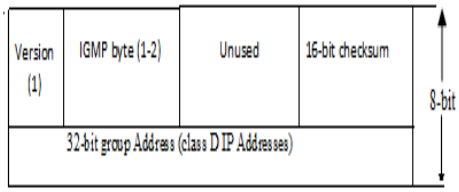


Fig 3: IGMP Packet.

IGMP consist of 32- bit as shown in the fig 3. it has the 32-bit group address of class D addresses. The class D is a higher – order four bits of 1110 and the range of IP Multicast group address is from 224.0.0.1 to address 239.255.255.255.

IV. IP MULTICAST GROUP

IP Multicast point to group of receiver points. IP multicast over ATM point to multipoint virtual circuits(VCs) as feature of dynamically. It creates the ATM point to multipoint is a switched virtual circuits gives the IP multicast traffic more efficient. Components required are in hardware are Computer network, switches, Ethernet cable, optical cable, and software are Linux based on Ubuntu OS (version 14.04), Wireshark analyser to analyses the real time results. VLC Player to streaming the video and audio of Multimedia files. Because of live program connection oriented network used (RTP) Real time transport protocol like audio and video of Live conferences. RTP is Monitor transmission statistics and quality of service(Qos) and synchronization of Multiple streams.

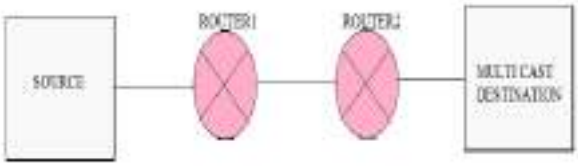


Fig 4: Block diagram of IP Multicast group.

The above fig 4 shows IP Multicast group of multimedia traffic. It consist of four network systems, first system is a server and two systems are act as a routers the last system has a client system or Multicast group. Router is used to forward the packets from the server(sources)to client(destination) network systems. For long distance communication I have used optical cable of 1Km for Multicast communication network. Multimedia is easy to watch and listen in the form of Text, audio, video. FM radio, this was shown by below fig 5 it has similar to the fig 4.

Fig 5: Multicast communication using 1KM distances of Optical cable.

V. ANALYSIS AND RESULTS OF REAL TIME APPLICATION

The above fig 5 Multicast communication of using Optical cable for 1KM distance.
Connecting the system network using switches by the desired Topology, configure the sever systems using linux based commands and also configure client systems. Similar to routers, then analysis the real time results of Audio, video. Comparison between the unicast and Multicast. Stored data is 2-3 Mbps in one second.

Fig 6: Multicast video file using 1km of optical cable.

The above graph of multicast video file consist along y-axis packets per interval and along x-axis time per second (.MP4).The videos are transmission of data from point to multipoint communication in computer network systems. it is one second of data is stored or frame of 2 Mbps.

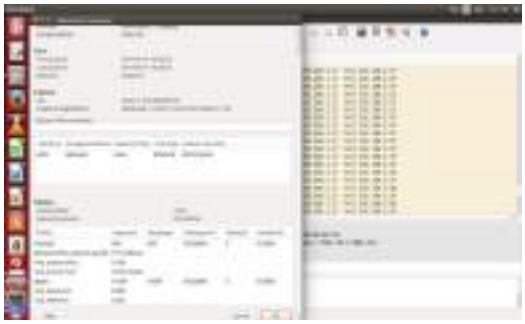


Fig 6: summary of Multicast video file via optical cable.

The above fig 6 shows the Multicast video via optical cable results in wirshark packets analyser.

Total number of packets are 284, first packet and last packet reached time is 1712.389 sec, averge per sec 0.166, average packets size 49.916 bytes, packets in bytes 284 *64= 14189.

Fig 7: Multicast video streaming file using optical cable.

A service where the data is delivery from sender to Multiple receiver network systems or Multicast groups has shown in fig5. it consist of graph of long Y-axis Packets per interval and long X-axis time per seconds. By connected the etho and eth1(Ethernet 1). Video file denoted as .Mp3. one second of data is stored in rang of 2-3Mbps of quality of real time result.

1KM distance. Multicast of traffic is less than Unicast communication network. Video file configure by access of live programs by capture the frames from the web camera.

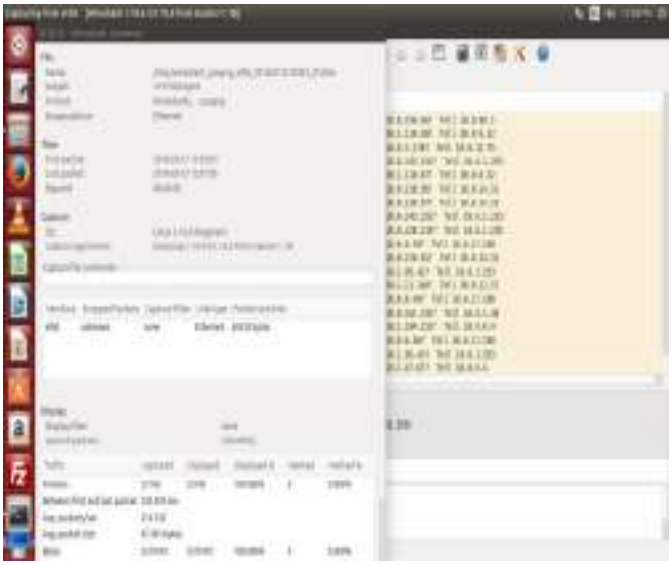


Fig 8. Summary of Multicast Packets for streaming process via fiber optics.

The summary of Multicast packets captured by 52166, Displayed 52166 and given packet displayed 100.00%, the average per sec is 241.742 average packet size 67.461bytes, Between first packet and last packet is 242.924 sec. Total number of packets in bytes written as 3519181.

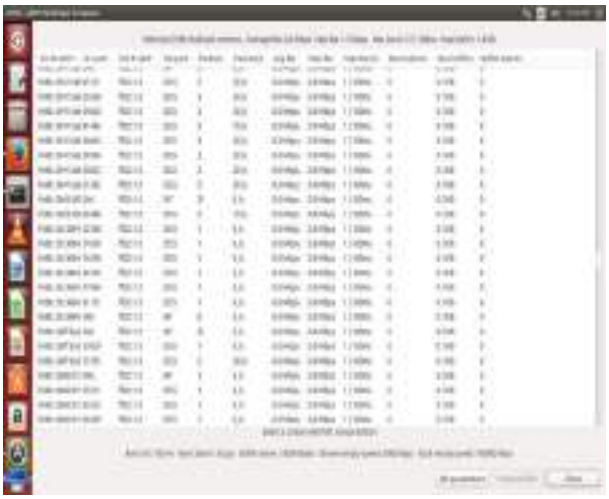


Fig 9.Multicast data streaming.

In this Multicast data streaming Eliminates the buffering while Transmission of live Audio or Video data from server to client bits Detected bits are 2988 Multicast streams. Max Bandwidth 1.5 Mbps and Max Buffer size is 1.4KB.

Multicast consist of IP address are IPV4 and IPV6 by given addresses.

The below table shows comparison of packets via fiber optics and streaming process.

Tab 1.1 The comparison of optical cable network and streaming of video.

Multicast video files(.Mp4)	To reach at the client system Time(sec) Packets	
1.Video for optical cable	1712.389	284
2.Video for optical cable by streaming process	242.924	52166

V. CONCLUSION

The analysis and Implementation of a Multicast group communication using optical cable 1KM based on real time transfer protocol presented in this paper. The multicast network is used in Live program communications like a cricket, live News, Audio, Video conferences without buffering the data. The main aim of this project is increases the life time of Network bandwidth and Avoids the delay than the Unicast communication. Achieve Wavelength is 1550nm and Bandwidth is 1.5Mbps. Unicast traffic is more than the Multicast communication.

VI. FUTURE SCOPE

The analysis is carried out Multicast communication better than the Unicast. In order to Improve the Network system or network communication. Network system proposed work is Broadcast communication. The data transfer from sender to all Receiver systems or many Network computer.

REFERENCES

[1] Adamson, Roca, "*Security and Reliable Multicast Transport Protocols Discussions and Guidelines",* Internet draft.

[2] Paul, S., Sabnani, K.k., lin. J,c., and Bhattacharyya, S."*Reliable Multicast Transport protocol(RMTP)*" IEEE journal on Selected Areas in communications,April 1997

[3] Peled and Vardi,"*Formal Techniques for Networked and Distributed systems*"

[4] C.Kenneth Miller," *Reliable Multicast Protocols and Applications"*, The Internet Protocol Journal –Volume 1,No.2

[5] Gemmel et al,"*The PGM Reliable Multicast Protocol"Wu, "Verification Analysis of Reliable Multicast Protocol"*

[6] Tanenbaum and Steen, "*Distributed system Principle and Paradigms"*

[7] *Internet Research Tsk Froce,Reliable Multicast Research Group* http://www.irtf.org/charter?gtype=old-rg&group=rmrg

[8] Adamson et al, "*NACK –Oriented Reliable Multicas Protocol (NORM)"*, RFC 3940

[9] Adamson et al, "*NACK-Oriented Reliable Multicast Protocol(NORM)",*Internet Draft-ietf-rmt-pi-norm-00.txt

[10] Speakman et al, *"PGM Reliable Transport Protocol Specification",* RFC 3208.

Clustering in Mobile Wireless Sensor Networks: A Review

Shantala Devi Patil, Vijayakumar B P
Dept.of Computer Science, REVA ITM, Dept. Of Information Science & Engineering, MSRIT
vijaykbp@yahoo.co.in, shantaladevipatil@gmail.com,

ABSTRACT- Wireless Sensor Networks (WSN) find its applicability from military to medical fields. Many of these applications requires node mobility. One of the systematic technique to incorporate efficient utilization of constrained resources in WSN is through clustering and also mobility aids to avoid hotspots. Clustering is a process in which the randomly deployed nodes of WSN are divided into non-overlapping groups called clusters and each one is assigned with a leader responsible for coordination and proper functioning of the cluster. We aim to provide a detailed insight to a novice researcher in this field with the state of art clustering schemes with their impact on the mobile WSN.

Keywords – Mobile WSN, Clustering, Node Mobility.

I. INTRODUCTION

Operations in Wireless Sensor Networks (WSN) must be resource efficient in order to maximize the lifetime of the network. WSN are networks with resource constraint and nodes are battery operated with no option of recharging. Clustering is one such operation to increase the longevity of the network. Clustering is a technique where nodes are divided into non overlapping groups, each group is governed by a node and will further be referred to as Cluster Head (CH) with all the nodes within a cluster referred as Cluster Members (CM). Though many clustering algorithms are proposed in literature for Ad-Hoc networks, these techniques are not applicable due to critical design goals of the WSN [22]. CH can be either pre-assigned by the network manager or it could be selected or elected by the other members in the cluster. The CH is assumed to be richer in resources as compared to that of the other nodes in Cluster. The CM are fixed for a static network or else if the nodes are mobile the CM keep changing.

Organization of the paper is as follows, in section II we discuss about the advantages of Clustering. Section III discussed the factors that influence clustering, followed by overview of clustering schemes based on mobility in section IV. In the last Section V we conclude with open research issues of clustering in WSN with mobile nodes.

II. CLUSTERING ADVANTAGES

2.1. The Network Model

A Mobile WSN is made of innumerable number of nodes that are deployed in hostile area where human reachability and operability is nearly impossible. Hence these nodes once deployed should self-organize and work appropriately to cater to the needs of the network. These nodes are battery operated and hence the lifetime of network is constrained. There are two types of nodes possible in WSN, static or mobile. Here in this paper we consider nodes are mobility, hence the name Mobile WSN. The main aim of any scheme should be to efficiently utilize the resources to maximize the network lifetime.

The Mobile WSN architecture [20] includes, Base Station (BS): The BS is the node with high energy responsible for processing and providing interaction interface between end user and other members, Clusters: Clustering is organizing the nodes of WSN into groups, Cluster Members (CM): are the nodes within a cluster responsible for sensing the environmental parameters and communicating to CH, they are capable of moving from one location to another. Cluster Head (CH): Leads the cluster members by coordinating cluster activities, aggregation of environmental data sensed by the CM, and they may also possess the capability to move. End Users: are the ones that use the data sensed and collected. They can either query or receive the reports from WSN.

2.2. Advantages of Clustering

Some of the advantages of clustering in WSN are enlisted below,

- Clustering supports scalability of WSN [19].
- Organization of nodes into clusters will localize the routing tables and also reduce the size of the routing tables stored at each node [22].
- The number of interactions and exchange of redundant messages among the nodes is reduced due to the organization of nodes into cluster [22].
- Cluster Heads evenly distribute the activities among the CM to balance the load and aid in increasing the performance of WSN [18].
- Clustering reduces the topology maintenance overhead by stabilizing network topology [19].
- Handling network is easier with clustering as Global changes made will not affect the networks locally [20].
- CH in order to prolong the life of network imposes strategies on its CM by scheduling the activities and manage whole of the cluster [22].
- CH aggregates data and passes this data to the BS hence reducing the number of message exchanges if all nodes had to communicate with the BS [21].

III. FACTORS INFLUENCING CLUSTERING

Clustering in resource constrained WSN should be energy efficient and hence it's important to consider the factors that influence such a process,

Node Type: Nodes in WSN are of two types Homogeneous and Heterogeneous. For Clustering with Homogeneous nodes, all nodes in the cluster have equal probability of being selected as the CH. The CH can conserve its energy by delegating the sensing work to the CM and perform only significant operations such as reporting to the BS or other CH, Data Aggregation and CM coordination. In case of clustering for Heterogeneous nodes [22], the nodes have different levels of energy. Nodes with highest level of energy perform the task of CH and nodes with low energy level perform the common task of sensing environment factors. Hence clustering of Heterogeneous nodes impose more constraints. [20,21,22] Shows that Heterogeneity of nodes is more appropriate for WSN.

Mobility of nodes: The nodes of the WSN can be of Static: the nodes are in fixed location till the network dies, Quazi: the nodes are static until the bootstrapping is complete then the nodes can continue to be static or move, Mobile: the nodes move from one location to another in the network. In such a network with mobility, clustering is very challenging due to network dynamics. Not much work is done in the clustering of nodes with quazi or mobile capability.

Mobility of Cluster Head: Clustering is affected if the CH is Mobile, as the cluster membership changes. In case of stationery CH, clustering and its management is very convenient.

Cluster Head Selection: When clustering is done with pre-assigned CH, the number of clusters will be fixed in the network and in case of elected CH then the number of clusters in the WSN varies. Many election algorithms are published in literature.

Network Topology: The organization of nodes in the network can be Flat or Hierarchical. In Flat topology, all nodes have similar function whereas in al topology nodes with different capabilities are present. Clustering is organizing the network into multiple hierarchy.

Clustering Process: The clustering process can be initiated by End User or BS and hence such a coordinated clustering is said to be centralized, whereas clustering without any central coordination or authority is said to be distributed. In Hybrid clustering process the CH election would be distributed and coordination within the cluster would be centralized and managed by the CH.

Number of Clusters formed: In some clustering schemes the number of clusters and the CH are Pre-defined and when the number of clusters in the WSN varies over time due to network dynamics then clustering is Adaptive.

Intra Cluster Topology: Clustering is easy task when the number of CM is Fixed and challenging task when the CM keep changing over time such an intra cluster topology is said to be Adaptive and very challenging for clustering.

Node Deployment: are of two types, 1: Deterministic - When the nodes are placed in pre-determined locations then clustering is preset. In Self Organized WSN the nodes are deployed randomly and makes clustering more challenging.

Connectivity between BS to CH: Single-Hop and Multi Hop.

Connectivity between CM to CH: Single-Hop and Multi Hop.

Convergence time of Clustering Process: In case of Constant convergence: the number of iterations for clustering process is constant regardless of the number of nodes in the network. Variable convergence is the one where the number of iterations for completing the clustering process is dependent on the number of nodes in the network at that instance.

IV. SOME OF THE PROMINENT CLUSTERING ALGORITHMS

From the literature, lot of work has been done in clustering of Mobile WSN. Mobility of nodes help in eliminating hotspots which are created when CH die due to excess load. These hotspots leave the network partitioned. Other advantages of mobility include scalability, increase in energy efficiency, improvement of network lifetime, and increased fault tolerance. Many schemes published in recent times are

Nasser et al [1] proposed a routing based on partitioning Mobile WSN into zones and a clustering with zone head election scheme. It has low overhead and is based on velocity of node, localization, mobility factor attributes. This scheme is applicable for nodes with low mobility.

Kim et al [2] proposed a scheme that is an extension of LEACH called LEACH-Mobile which included mobility. A node with low mobility is elected as CH and distributed evenly over the sensing area, once elected CH is stationary. CH is selected based residual energy, mobility and location attributes. A mobile node requests for joining new CH when it does not receive request from its CH for 2 consecutive data frames.

Santhosh et al [3] LEACH-ME scheme is an extension of LEACH and LEACH-M to support mobility. It uses remoteness concept for CH election with mobility, residual energy and GPS based location detection mechanism as clustering attributes. The CH being mobile but fixed BS. Node will wait for two message transmission failures before joining new CH.

Awaad et al [4] proposed a clustering scheme based on mobility and adaptability for mobile heterogeneous WSN called Clustering Based Routing Protocol (CBRP). Based on node residual energy, CH is elected. The nodes moved out of one cluster to another have to be reassigned to new CH. This scheme reduces packet loss and energy consumption.

Deng et al [5] proposed a Mobility based clustering (MBC) scheme with mobile WSN and CH selection is based on mobility and residual energy attribute. BS is fixed, nodes are mobile and homogeneous. The authors claim that the scheme is proactive with reduced control overhead, good packet delivery date but does not provide fault tolerance.

Sahi et al [6] proposed Enhanced MBC an extension of [5] which is reliable and fault tolerant in mobile WSN

environment. SN conserve energy and show its existence by sending special packets to CH when they have no data for transmission. The scheme is more energy efficient and increases network lifetime when compare to [5].

Anita et al [7] proposed an Enhanced CBR (ECBR) for Mobile WSN. It is an extension of [4] where CH election is based on high residual energy, low mobility, location awareness, distance between BS and node. The scheme is energy efficient where nodes are mobile and BS is fixed.

Jose et al [8] proposed a Mobile sink Assisted Algorithm (MSA). Both static and mobile sink are part of Mobile WSN with static sink controlling and coordinating the mobile sinks. The Clustering is based in LEACH based on residual energy and general behavior of nodes, along with path planning for mobile sinks. The mobile sinks collect that from various parts of network and pass this data to static sink for processing.

Zhao et al [9] proposed a Load Balanced Clustering and Dual Data Uploading (LBC-DDU) which divided the heterogeneous WSN in 3 layers Sensor Nodes Layer : is used to sense and collect the data, Cluster Head Layer: Multiple CH are selected for each cluster which guarantee the communication and connectivity. Mobile Collector Layer: the data aggregated by the second layer from first layer is collected by third layer called Mobile Collector Layer.

Zahhad et al [10] The Mobile Sink – Based Adaptive Immune Energy Efficient Clustering Protocol (MSIEEP) is based on Adaptive Immune Algorithm to find the stability period of the network and also optimal number of CH. AIA acts as a guide to mobile sink. This protocol improves lifetime and stability periods of heterogeneous network in mobile scenario.

Srithar et al [11] proposed an Adaptive Weighted Fuzzy Clustering Based Cluster Head Selection Algorithm (AWFCA) which tries to combat the problem of nodes distribution, holes in network and flat structure by incorporating heterogeneous mobile nodes. This scheme selects a representative path to CH based on weights using a centralized approach. The network lifetime is increased based on three fuzzy parameters, energy concentration and centrality.

Jason Li et al [12] proposed a Distributed Efficient Clustering Approach (DECA) a distributed clustering approach and is efficient with mobility of nodes. The clustering in DECA is based in the score function which is in-turn a calculation of residual energy of node, node computation capability, mobility, identifier and connectivity. The higher the score, more the chances of that node being selected as a CH.

Ali et al [13] proposed a scheme for distributed clustering for mobile WSN called Distributed Efficient Multi-Hop Clustering (DEMC). The mobility model for nodes in this scheme is mass mobility model, nodes are aware of their location and homogeneous in nature. The clustering involves recovery mechanism that reduces the packet loss during inter-cluster communication also improves inter-cluster communication between the CHs.

C Tang et al [14] This Collaborative Weighted Clustering Algorithm (CWCA) is based on weighted clustering using multiple metrics to calculate weight such as degree, Euclidean distance, relative mobility, lifetime of a node. This scheme has dual applicability to both static and mobile sensors. Some of the potential benefits of the scheme are energy efficiency, higher power gain and distributed sensing.

S. Park et al [15] proposed a Two Tier Clustering (TTC) scheme for mobile environment based the self-organizing capability. The clustering is done in two phases which make sure that the CHs are distributed evenly in the mobile network and also prevents the CH clumping. TTC is able to reduce data and also provide high energy efficiency.

S Basagni [16] proposed a Generalized Distributed Mobility Adaptive Clustering (G-DMAC). This scheme requires each node to have local knowledge and direct communication with the CH, this fastens the intra and inter communications between the nodes. Cluster formations is based on weight and mobility parameters, the nodes are mobile even during the clustering set up. The stability of this clustering scheme is measured in the terms of number of elections and affiliations.

V. CONCLUSION AND RESEARCH ISSUES

Mobility of nodes in WSN helps achieve scalability, fault tolerance and also energy efficiency. From the literature survey it is very evident that not much work is carried out about clustering nodes with mobility. And those schemes available are not optimal as they do not consider all the issues.

Node Mobility factor is not appropriately addressed and majority of the clustering schemes are based on the static nodes in the network or just considering the mobility of the CH. These conventional schemes proposed for static WSN find it hard to deal with mobile WSN.

5.1. Other Issues of the published work include:

Even if the mobility of nodes is considered in a scheme, the work does not focus on the network dynamics, where nodes migrate to different clusters. The joining and leaving the cluster procedure has to be properly addressed.

When considering securing the WSN with mobility almost no work or very insignificant work is carried out. No proper schemes are present to address the dynamic network management, Network adaptability. The current performance metrics used for the static WSN are not suitable for the mobile WSN, hence new metrics are required to measure the performance such networks. Other design issues that still need to be addressed in mobile WSN are CH or nodes moving out of communication range and appropriate localization and security schemes.

The Clustering process can be initiated in the network based on need i.e on demand to conserve resources. This issue is still unaddressed.

The paper provides an insight regarding clustering in WSN with nodes mobility summarized in Table 1 and also presents the scope for future work for the same.

REFERENCES

[1] Nasser, N., Al-Yatama, A., Saleh, K,"Zone based routing protocol with mobility consideration for wireless sensor networks" Springer, 2012.
[2] D.S. Kim, Y.J. Chung,"Self-organization routing protocol supporting mobile nodes for wireless sensor network," Proc. First Int. Multi-Symp. on Computer and Computational Sciences, Hangzhou, China, pp. 622– 626, 2006.
[3] G Santhosh kumar, Vinn Paul M V, K Poulose Jacob,"Mobility Metric based LEACH-Mobile protocol", ADCOM 2008 IEEE.
[4] S.A.B. Awaad, C. K. Ng, Mohd F.R.,"Cluster based routing protocol for mobile nodes in wireless sensor network," Proc. Third Int.Conf. on Quality of Service in Heterogeneous Wired/Wireless Networks, Waterloo, Ontario, Canada, pp. 233–241,2006.
[5] Deng, S., Lil, J., Shen, L., "Mobility-based clustering protocol for wireless sensor networks with mobile nodes," IET Wireless Sensor Systems, Vol. 1, Iss 1, pp.39–47, 2011.
[6] Lofty Sahi, Padmavathi Khandnor,"Enhanced Mobility Based Clustering Protocol for Wireless Sensor Networks", International Conference on Electrical, Electronics and Optimization Technique ICEEOT-2016.
[7] R.U.Anitha, P.Kamalakkannan.: 'Enhanced Cluster Based Routing Protocol for Mobile nodes in Wireless Sensor Network'.Proc. of the 2013 IEEE Int. Conf. on Pattern Recognition, Informatics and Mobile Engineering, Feb, 2013, pp. 187-193.
[8] Deepa V Jose, Dr G Sadashivappa,"Mobile Sink Assisted energy efficient routing algorithm for wireless sensor networks", World of Computer Science and Information Technology Journal (WCSIT) Vol 5, No 2, 16-22, 2015.
[9] M Zhao, Y Yang, C Wang," Mobile Data Gathering with Load Balanced Clustering and Dual Data Uploading in Wireless Sensor Networks", IEEE Transactions on Mobile Computing, Vol 14 Issue 4, April 2015, 770-785.
[10] M Abo-Zahhad, S M Ahmed, N Sabor, S Sasaki,"Mobile Sink Based Adaptive Immune Energy Efficient Clustering Protocols for improving the lifetime and stability period of WSN", IEEE Sensors Journal Vol 15, Issue 8, August 2015.
[11] S. Srithar, Dr K Karuppasamy, "Energy Efficient daptive Weighted Fuzzy Clustering Based Routing Protocol in Sensor Networks", International Journal of Engineering Research and General Science Vol 3,Issue 6, Nov-Dec 2015.
[12] Jason H Li, Renato Levy, Miao Yu,"Mobility Resistant Clustering in Multi-Hop Wireless Networks ", Journal of Networks, Vol 1, No 1, May 2016.
[13] Shahzad Ali and Sjjad Madani, "Distributed Efficient Multi hop clustering protocol for mobile sensor networks", The International Arab Journal of Information Technology, Vol 8 , No 3, july 2011.
[14] Chengpei Tang, Sanesy Kumcr Shokl, George Modhawar, Qiang Wang,"An Effective Collaboration Schemes for weighted Clustering Schemes for Energy Balancing in Wireless Sensor Networks", Sensors 2016, 16(2), 261.
[15] Sun-Young Park,"Enhanced Self Organization Clustering in Mobile WSN ", Advanced Science and Technology Letters, Vol 51, CES Cube 2014, PP 10-14.
[16] Basagni, Stefano, "Distributed clustering for ad hoc networks." Parallel Architectures, Algorithms, and Networks, 1999.(I-SPAN'99) Proceedings, Fourth International Symposium on. IEEE, 1999
[17]. Garcia-Hernandez et al, "Wireless Sensor Networks and Applications: A Survey", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.3, March 2007.
[18]. Dr Sami Halawani, Abdul Waheed Khan, "Sensors Lifetime enhancement Techniques in Wireless Sensor Networks: A Survey", Journal of Computing, Vol 2, Issue 5, May 2010.
[19]. Akyildiz I F, Y Sankarasubramaniam, E. Cayirci,"Wireless Sensor Networks: A Survey", Elsevier, Computer Networks 38 (2002) 393-422.
[20]. Ameer Ahmed Abbasi and Mohamed Younis,"A Survey on Clustering algorithms for Wireless Sensor Networks", ELSEVIER Computer Communications 30, 2007.
[21]. Mohamed Younis and Kemal Akkaya, "Strategies and Techniques for node placement in Wireless Sensor Networks: A Survey", ELSEVIER, Adhoc Networks 6, 2008.
[22]. Sanjeev Kumar Gupta, Neeraj Jain and Poonam Sinha, "Clustering Protocols in Wireless Sensor Networks: A Survey", International Journal of Applied Information Systems, Vol 5, No 2, 2013.

Field weakening strategy in vector control of induction motor during voltage saturation region

Pallavi H.R., Sumitha T.L.
Department of EEE, T.O.C.E, Bangalore
pallavihr62@gmail.com, sumitha.jaiber@gmail.com

ABSTRACT - The main aim of the project is to obtain field oriented vector control of induction motor during the inverter saturation. From this vector control method, it is possible to utilize the maximum output voltage available from the inverter completely without exceeding the voltage and current rating of controllers. When the inverter voltage gets saturated, the quadrature axis current controller modifies the flux reference instead of directly obtaining the quadrature axis voltage reference, because when the inverter voltage saturates, there is one degree of freedom available, i.e., either the direct axis voltage or quadrature axis voltage can be controlled independently. Space vector switching strategy is used to control the inverter gate. Hence low switching frequency operation of inverter is possible, which makes it attractive scheme for the traction system. MATLAB based implementation is carried out for this project. Tools used are FPGA controller, Induction motor, Interfacing board, 2-level inverter

Keywords - Field weakening, FPGA Controller, Over modulation, Vector control , Voltage Saturation

List of Symbols

σ	Leakage Co-efficient
L_m	Magnetizing inductance
i^*_{mr}	Flux current reference in rotor flux oriented squirrel cage induction motor drive
ω_s	Stator frequency of motor in $rad/_{sec}$
ω_m	Mechanical frequency of motor in $rad/_{sec}$
ω^*_m	Reference mechanical frequency of motor
$i^{act}_{sR}, i^{act}_{sY}, i^{act}_{sB}$	Sensed instantaneous stator currents of RYB phases respectively
$i^{fun}_{sR}, i^{fun}_{sY}, i^{fun}_{sB}$	Estimated instantaneous fundamental stator current of RYB phases respectively
r, s	Represents rotor and stator respectively
$V^{ref}_{sR}, V^{ref}_{sY}, V^{ref}_{sB}$	Instantaneous stator voltage references of RYB phases respectively
V^*_{sd}, V^*_{sq}	d-axis and q-axis stator voltage reference respectively
V^{max}_s	Maximum length of stator voltage space vector = $\sqrt{V^2_{sd} + V^2_{sq}}$
i^{max}_s	Maximum length of stator current space vector = $\sqrt{i^2_{sd} + i^2_{sq}}$
i^*_{sd}, i^*_{sq}	Stator current reference for d-axis and q-axis

I. INTRODUCTION

The variable speed drives are widely used in industrial application. Most of industries use AC motors for their processing. The standard in those drives are induction motor and synchronous motors. The induction motor control has two types of speed control that is scalar control and the other is the vector control. V/F control is an example of scalar control and Indirect Field Oriented Control and Direct Field Oriented Control comes under vector control techniques. In this project direct vector control method of Vector Control is used.

Field Weakening stage of operation in induction motor is a recently introduced method where the control of induction can happen during the saturation of the inverter voltage. This field weakening stage occurs at the place where the voltage saturation occurs in the inverter that is when the inverter operates at the six-step mode of operation.

II. BASIC BLOCK DIAGRAM

The Basic block diagram is as shown Fig.1 .It contains mainly two parts-The induction motor fed from 2-level inverters and FPGA controller. The line currents of induction motor and D.C. input of the inverter are sensed from the current and voltage sensors respectively. The interfacing board converts the sensed physical quantities into a range of A/D converters. Thus obtained input from the interfacing board, A/D converters the physical quantities into digital quantity and gives the output to the Field programmable gate array(FPGA) controller.

In FPGA there are mainly four units-SVPWM switching, Speed estimation block, Vector control with flux weakening and Harmonic current estimation. This proposed method uses sensorless speed estimation method. The speed of rotor can be estimated from speed of stator and speed of air gap as explained in[1]. The line currents of induction motor contains harmonic, so it cannot be used with the low frequency switching (IGBT switch) which are used in

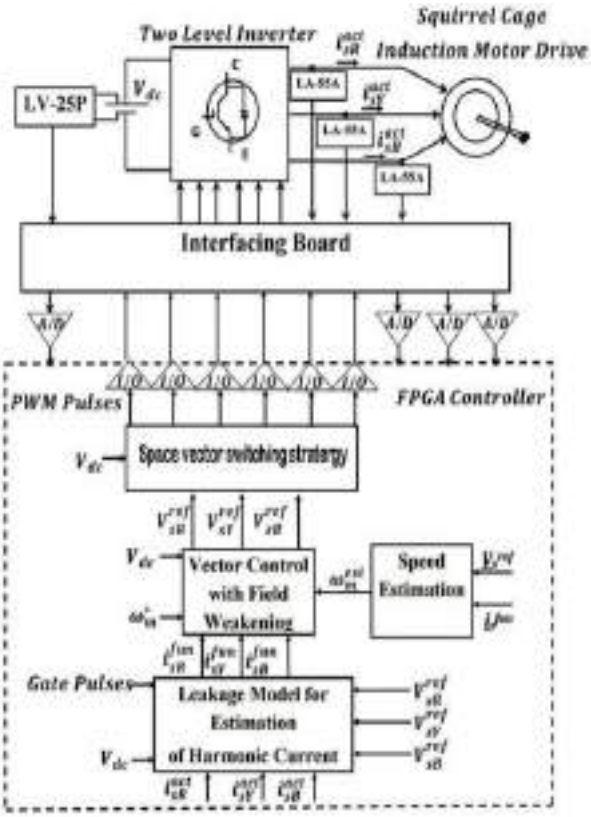


Fig.1.Basic Vector Control diagram

inverter. Filters cannot be used to remove the harmonic current as it hampers the current controller. So the harmonic currents also need to be estimated as explained in [2].Thus obtained fundamental line currents are fed to the vector control unit. The main focus of project lies on vector control with flux weakening. The flux weakening, an integral part of vector controller, enables the use of maximum output of inverter and also avoids the current controller from the saturation. Input pins of FPGA gives the gating pulses to the gate drivers of switches through the interfacing board. Interfacing board increases the gate pulse voltage to 15V D.C.

III. MODIFICATION IN VECTOR CONTROL

The conventional basic block diagram for field oriented control system is as shown in Fig.2 .This block diagram is sufficient to run the motor below the base speed in rotor flux oriented reference frame.

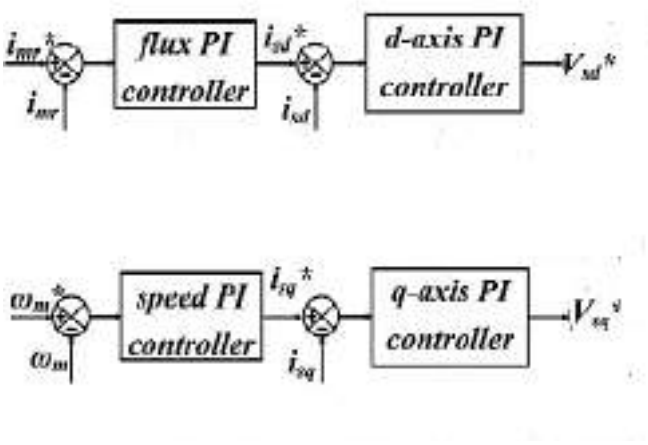


Fig. 2. Current controller in Vector diagram

The conventional basic block diagram for field oriented control system is as shown in Fig.2 .This block diagram is sufficient to run the motor below the base speed in rotor flux oriented reference frame.

Modification of conventional vector control to operate above the base speed is as shown in Fig.3. . Flux current reference i^*_{mr} and torque current reference i^*_{sq} are needed to be modified in order to operate in field weakening above the base speed. The block FF i^*_{mr} block provides the flux current reference and the block FF i^*_{sq} provides the torque current reference to operate the motor above the base speed.

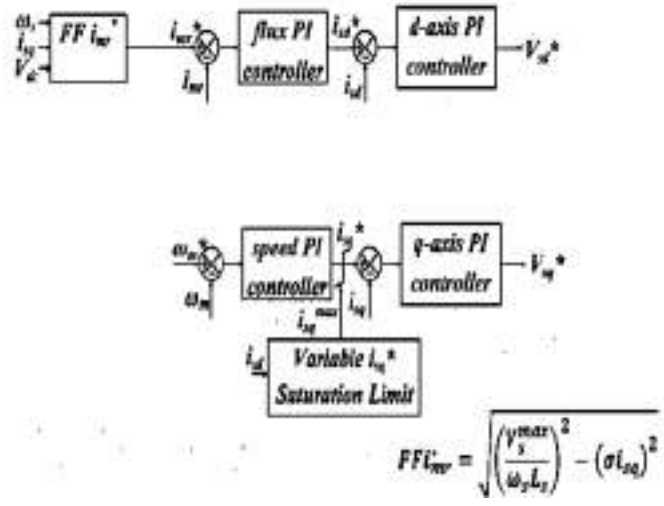


Fig. 3.Modification in current controllers of vector diagram

1 Modification in flux reference

As the speed increases, the flux gets reduced. In order to keep the flux to the optimum value flux current reference is calculated from the below equation.

$$FF\ i^*_{mr} = \sqrt{\left(\frac{V_s^{max}}{\omega_s L_s}\right)^2 - \left(i_{sq}\sigma\right)^2} \quad (1)$$

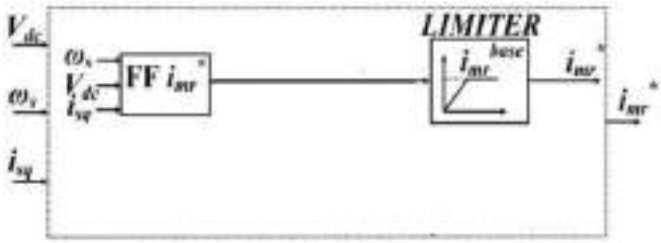


Fig.4. Modification in flux current

In order to utilize the maximum output from inverter, stator voltage is kept at the maximum value which can be obtained from state stator voltage equation [3].Below the base speed, the flux reference is very large. Therefore the flux reference value is fixed within a limit as shown in the Fig. 4.. Below the base speed, flux value is kept at the rated value.

2 Modification in torque reference

In order to have maximum torque for a given speed reference the torque reference i_{sq} is kept at its maximum value without making the current controller to saturate. Therefore the torque current reference is calculated from equation (2).

$$i_{sq}^{max} = \sqrt{(i_s^{max})^2 - i_{sd}^{\ 2}}. \quad (2)$$

Field weakening region is split into two regions- flux weakening region 1and flux weakening region 2. Flux weakening region 1lies below the constant slip region and rest of region stay in flux weakening region 2.In flux weakening region 1 the flux current reference is calculated from $i_{sq}^{max} = {}^{i_{sd}}/_{\sigma}$ is greater than the value of i_{sq}^{max} calculated from $\sqrt{(i_s^{max})^2 - (i_{sd})^2}$ which in turn helpful to find flux weakening region 1 or 2. The flux reference is altered from i_{sq}^{*} to i_{sq}^{max} helpful to run the motor at maximum possible torque [3].

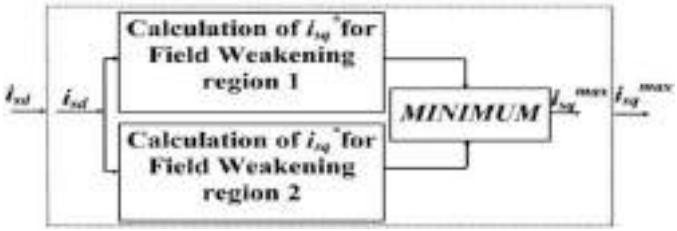


Fig.5.Modification in Torque reference

Modification in quadrature and direct axis voltages-Quadrature axis voltage (V_{sq}^{*}) and direct axis voltage (V_{sd}^{*}) are obtained from the quadrature and direct axis current controllers respectively. Above the six step mode of inverter, it is not able to control the Induction motor as it is impractical. If the voltage saturate, either the V_{sd}^{*} or V_{sq}^{*} should be varied to bring the inverter voltage within the six step mode. Therefore a limiter is used to compare the voltage of V_{sq}^{*} and $\sqrt{(V_s^{max})^2 - (V_{sd})^2}$.

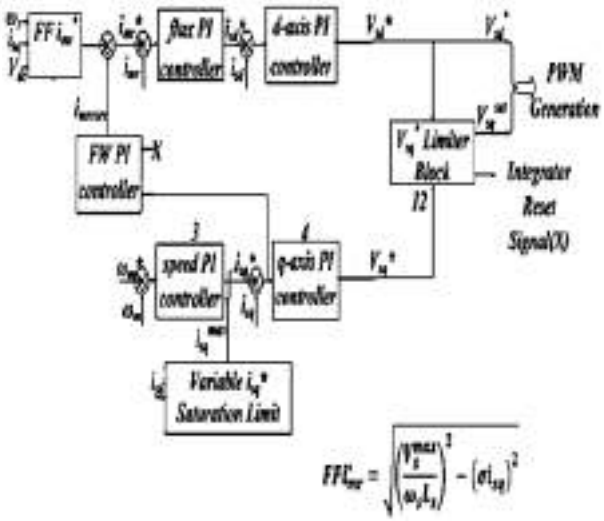


Fig.6. Modification in current controller

If V_{sq}^{*} obtained is greater or equal than the$\sqrt{(V_s^{max})^2 - (V_{sd})^2}$, then onwards the quadrature axis voltage is varied from the flux current (i_{mr}^{*}) or from the i_{sd}^{*}. Since variation in i_{sd}^{*} leads to variation ini_{mr}^{*}, it is preferable to control the V_{sq}^{*} voltage from the i_{mr}^{*} still the inverter output voltage becomes lesser than the six step mode. The signal X integrator is used to check whether the quadrature axis voltage is within the limit and if the quadrature axis voltage is exceeded the signal X sends the signal to FW PI controller as shown in Fig.6.

The signal X is generated form the above Fig 7.. The current controller $i_{sq}$is always active, it remains in the saturated only when the six step mode of operation.

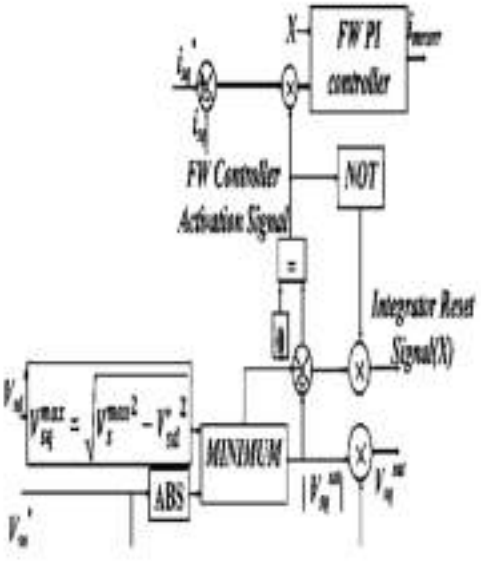


Fig.7. Generation of signal X

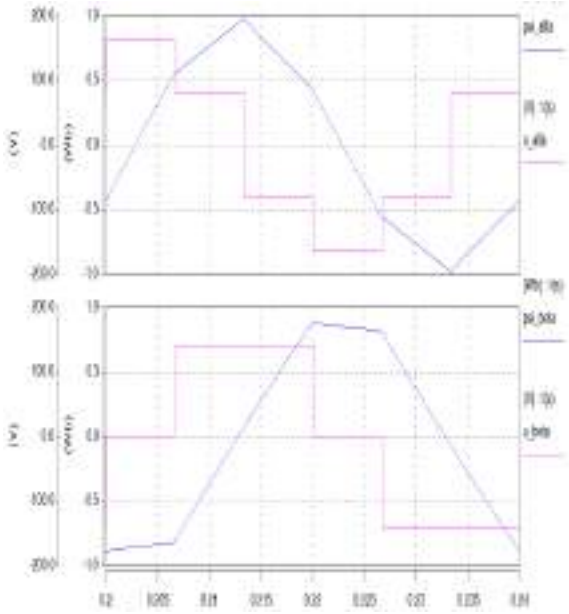


Fig.8.Simulation of inverter under the six step mode operation

IV. CONCLUSION

This paper proposes, the modification in flux weakening strategy when the inverter voltage enters to the six step mode of operation. From the modified method of flux weakening it is possible to current controller even at inverter saturation method. In this paper, space vector strategy which can be extended in over modulation region for low switching and fundamental line current of induction motor from harmonic current estimation is done.

REFERENCES

[1] Saroj Kumar Sahool, Tanmaoy Bhattacharya,*A synchronized sinusoidal PWM based Rotor Flux Oriented controlled induction motor drive for Traction application*, Depart.of electrical engineering,IIT Kharagpur, ,saroj@ee.iitkgp.ernet.in, btanmoy@ee.iitkgp.ernet.in2

[2] B.J.Siebel , T.M. Rowan ,R.J.Kerkman, member,*Field-oriented control of an induction machine in the Field weakening region with DC-Link and load Disturbance Rejection* ,,IEEE1,member,IEEE2,senior member,IEEE3.

[3] S.H.Kim,S.K.Sul, *Maximum Torque control of an induction machine in the Field weakening region*,,member IEEE

[4] W.Leonhard,*Control of Electrical Drives*

[5] R.Krishnan,, *Electric Motor drives, Modelling, analysis and Control*

BIOGRAPHIES

Pallavi H.R. is born in , 1991.She recived the B.E. degree in electrical & electronics from VTU, Belagaum. She is presently doing the M.Tech. in industrial drives.

Investigating Hackers on Facebook Application using FRAppE

Asha Alias[1], Rincy Varghese[2], Ritu. M. Varghese[3]
[12&3]Department of Computer Science, visvesvaraya Technological University, Belgavi, Karnataka
Email: [1]ashaalias95@gmail.com, [2]rincyvgs@gmail.com , [3]ritu_varghese@yahoo.com
[*4]**Mr. V.M. Saravana Perumal**
[4]Assistant professor, Department of Computer Science, visvesvaraya Technological University, Belgavi, Karnataka
Email: [4]saran4umohan@gmail.com

ABSTRACT-There are millions of people who install the facebook application and third party apps are always the major problem for fame and addictiveness of Facebook. Thus hackers have known the strength of the apps for spreading the unwanted things. We have found that 15% of apps are malicious. Thus for the problem we have developed our contribution ....i.e. FRAppE (Facebook Rigorous Application Evaluator).FRAppE is to focus on the investigation of malicious apps on facebook.The behavioral sense of 111k Facebook apps in 2.6 million users were used to develop FRAppE. With 99.5% accuracy FRAppE can detect malicious apps with no false positives and a low false negative rate (4.1%).We have a group of features to identify the spam app from good apps.

For app testing and ranking we see FRAppE as a step towards creating an independent watchdog that warn users on Facebook before installing apps.

Keywords –Benign apps, Facebook applications, Malicious apps, MyPageKeeper

I. INTRODUCTION

One of the most popular application which comes with its own advantages and disadvantages is Facebook. Such enhancement consist of interesting and enjoyable ways of communicating among online friends and it also include interesting games and listening to music .Now a days we can see that there are 500k apps are available on Facebook ,within that 40M apps [1]are installed everyday by the Facebook users. In addition many apps get acquired and maintain a sizable user.

Unfortunately recent evidence shows that, hackers have started deploying malicious apps [7, 9] can provide a lucrative business for hackers. Hackers can benefit from a malicious app in many ways. i) The app can obtain users personal information including password, email id, gender .ii)The app can spread spam in a large number of users. Here
the problem is, there are many malicious apps spreading on Facebook every day[6].

Today, the user has very limited information about the apps at the time of installing it on Facebook. That app may be malicious. This is an open gate for the hackers to obtain the personal information from users.

To protect the Facebook users from hackers, we develop FRAppE, a suite of efficient classification technique for identifying whether an app is malicious or not. To develop FRAppE, we use data available from MyPageKeeper, is a Facebook app [36] designed for detecting the malicious posts on Facebook. That will check the Facebook profiles of 2.2 million users. FRAppE (Facebook Rigorous Application Evaluator) is a tool which is mainly focused on detecting malicious apps on Facebook. It is an effective detection approach.
Following are our key contributions.

- *FRAppE can provide 99% accuracy in detecting malicious apps.*

We build FRAppE to detect the malicious app on Facebook using on-demand and aggregation based app information. By adding aggregation based information, FRAppE can discover malware apps with 99.5% accuracy with no false positive and lower false negative(4.1%).

- *The profile of spam and good apps is different.*

The malicious app profiles are significantly different from those of benign apps. Most of the malicious app have the same name. The benign app, that provides similar functionality.

- *15% of apps on facebook are detected as malicious.*

The evidence shows that around 15% of apps on Facebook are malicious .And 100k users each by convincing them to follow the links on the posts made by these apps.

II. OVERVIEW

2.1 Apps in facebook

Third party apps developers have rights in Facebook to offer services to the user. If the user installs the Facebook application to his profile ,the user allows the application server to access the permission to a set of details that the user have provide in his Facebook profile like email address and also permission to access the some action in behalf of user such as post on the wall. By handling O Auth 2.0 token, Facebook allows this permission to any applications, this token is allocated for each user who installs the facebook.fig 2.1 show how hackers make use of the malicious apps, the malicious apps are works as follows:

- Hackers promote the user to install the apps by giving some false rewards with some keyword "Free", "Real", "Hurry".

- After installing the app it will provide the user a new web page where the users need to give some action regarding that reward such as complete task with false promises again.
- Then it will ask for personal information from profile.
- After that the app makes malicious post on this user's wall.

Figure 2.1 system design

2.2 MyPageKeeper.

MyPageKeeper[36] is a security app provided by facebook application. This MyPageKeeper discover the malicious posts on the the user's wall then apply url blacklisted as well as SVM classification technique to detect malicious apps. figure 2.2 shows the architecture design of the FRAppE.In existing system MyPageKeeper discovers only post of hackers with 97% high percentages of accuracy [28].

MyPageKeeper used Support Vector Machine (SVM) based classifier to discover whether the URL is malicious or benign. The classifier identifies the malicious post by taking some features consist of the presence of some keywords such as "click here" ,"free", and "fast" and also by the resemblance of text messages and number of the likes and comments if the level of likes are lower than it is malicious. If the URL is found as malicious the all the post contains in that URL will be malicious,

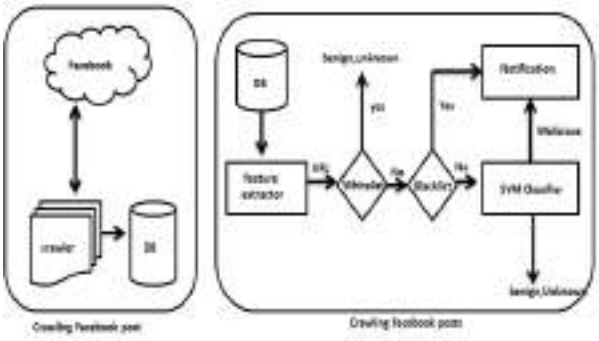


Fig 2.2 system architecture

2.3 Dataset

Over 2.2 millions install Facebook a day so Facebook apps have dataset from 2.2M Facebook user. Which has be followed by MyPageKeeper. This dataset consist of 124M posts from 2.2M walls which followed by by MyPageKeeper [13]. By investigating the faceboook, Post over 9 months from June 2013 to March 2014. This 124M post is made by 111K apps.

In the investigation we have to give some sample dataset they are by:

- Discovering malicious applications if any post has found as malicious so the application which that post has made can mark as malicious post[6].

  In several investigating we found "Death predictor" user also marked as malicious . this use already describes that addictiveness of facebook users. To prevent those kind of misuses, we used whitelist to classify the benign apps from URL. After whitelisted we left 6,273 malicious applications
- We also investigated about apps permission to be granted inorder to installs the application.All the application which are licensed is provided with an app_id(*httpd://www.facebook.com/apps/graph_apps?id=app_id).*By crawling all the apps this URL has been checked and detect the benign and malicious apps.

III. WIDESPREAD OF MALICIOUS APPS

The factor for identifying malicious apps and the main reason of it is that the malicious pots are posted by these apps on facebook.53% of malicious posts by MyPageKeeper was posted by malicious apps.

There are two different ways of widespread of malicious apps:-

(i) 100 thousand clicks on the URL's posted are got by 60% of malicious apps.

We determine the number of clicks for malicious apps on the links which are the malicious post .For the malicious apps in D-Sample dataset, we reach all bit-ly URL's in posts. We observe more onbit-ly UPL's since bit-ly offers an API[18] for receiving number of clicks is lower bound.

Even bit-ly link will receive clicks from various other sources outside facebook.For this purpose for the total number of clicks received in bit-ly URL ,is an upper bound and it is done through facebook. Almost 6,273 malicious apps in D-Sample dataset, it is known that 3,805 of the apps has posted 5,700 bit-ly URL's in total.

We usually observe and query bit-ly for the click count in each URL.60% of malicious apps cover over 100k clicks in which 1M clicks are received by 20% each,The most eye catching was the one with 1.742,359 clicks i.e :What is the sexiest thing about you?".

(ii)There is a median of 40% malicious apps with 1000 monthly active users.

By inspecting number of users on facebook we examine the malicious apps.In order for the above study we use Monthly Active Users (MAU) provided by facebook for every app. We found that 40% of malicious application had a median MAU of atleast 1000 users, and 60% of malicious application achieved 1000 during three month observation.

One of which it became famous was "Future Teller" which had maximum MAU of 260,000 and median of 20,000.

a. Posting direct links to other apps.

We find 692 promoter apps in our D-Sample dataset which promoted 1,806 different apps using direct links. The activity was intense :15% of the promoters promoted atleast 5 promote apps .For example, 'The App' was

promoting 24 other apps with names 'The App' or 'La App'.

b. Indirect app promotion:

Hackers have started using websites outside facebook to have promotion of apps. We can know the malicious apps as they contain shortened URL. If the problem from URL is identified and solved it directly points to the other website forward users to different app installation pages.

IV. PROBLEM DEFINITION

From our observations we find that malicious app are on Facebook. Our next step is to build a tool that must identify malicious content on Facebook. To develop a tool like FRAppE, we should analyze and compare the various features of malicious and benign apps. There are two divisions of features: on-demand features and aggregation based features.

4.1 On-demand features

The on-demand features comes with an application, which tells that one can obtain the on-demand feature given the application's ID. such metrics consist of name of the app, description, company, category and permissions.

4.1.1 Application summary

Malicious apps have incomplete application summary. In the first step, we compare malicious and benign apps with respect to application present in the application's summary such as app description, company name and category. Only 1.4% of malicious apps have a non empty description, whereas 93% of benign app configures their summary with a description.

4.1.2 *Required* permission set

97% of malicious apps require only one permission from users. Every Facebook application requires the authorization from the user before using it. And every app requests the user to provide the set of permission at time of installation. These permissions are chosen from a pool of 64 permissions pre-defined by Facebook.

4.1.3 Redirect URI

Malicious app redirect user to domains with poor reputation. In an application's installation URL, the redirect URL parameter refers to the URL where the user is redirected to once she/he installs the app. We extracted the redirect URI parameter from the installation URL for apps in the D-Inst dataset and queried the trust reputation score for these URIs from WOT [8].

4.2 Aggregation-based features

Now, we analyze applications with respect to aggregation -based features. Unlike the features we considered in on -demand features. we considered so far, aggregation based features for an app cannot be obtained on-demand. Here we envision that aggregation -based features are assembled by entities that will check the posting behavior of various application across users.

4.2.1 App name

85% of malicious apps have an app name identical to that of at least one other malicious app. An application's name is fixed by the app developer at the time of the creation of that app .And every app has a unique app ID, Facebook does not impose any restrictions on app names. So it is possible to create multiple apps with the same app name.

4.2.2 External link to post ratio

Malicious app often post links pointing to domains outside Facebook , whereas benign apps rarely do so. Every post on facebook include an URL. These URL may be made by malicious or benign apps. We can see that 80% of benign apps do not post any external links, whereas 40% of the malicious apps have one external link on average per post. This shows that malicious apps attempt to lead users to web pages hosted outside facebook, whereas the links posted by benign apps are almost always restricted to URLs in the facebook.com domain.

V. INVESTIGATING HACKERS ON FACEBOOK

We have classified the hackers apps which is malicious and benign apps, we have 2 variants to this classifier they are FRAppE lite and FRAppE. The security apps of Facebook that is MyPageKeeper only discover the malicious post and links but not the apps. These two variants of classifier is designed to discover the malicious apps.

5.1 FRAppE lite

This lightweight version will only make use the application feature of On-Demand. On-Demand specifies with respect to the app_id and FRAppE lite crawls the application with respect to these On-Demand features.

We use SVM [15] classifier to classify the hackers and benign. The FRAppE lite will be giving the accuracy 99.0%, with low false positive (0.1%) and false negative(4.4%) accuracy is defined as the ratio of truly identified apps which benign or malicious , false positive rate is fraction of benign apps incorrectly as malicious.

5.2 FRAppE

There are 2 features used to classified the malicious apps and benign apps, this FRAppE uses the aggregation based features with the On-Demand features that it's lightweight version only uses the On-Demand feature. Aggregation based feature of an app which consist a cross user and cross-app view with time.

FRAppE which gives the accuracy with 99.5% and with 4.1% of false negative rate also it doesn't contain any false positive. We invent FRAppE which is used in Facebook and also secure from third party application of millions of users

5.3 Ways to discuss New Hackers

We used to crawl all the posts, links and apps in the user's wall to do so we apply FRAppE to all URLs. If any new apps has discovered it will discover the malicious URL by using different ways they are

1. Facebook used to keep checking the hackers in Facebook application then it discover and disables from the wall by using the graph which contains the malicious app list. This has done by API in Facebook (https://graph.facebook.com /appId) which returns false for a malicious app because its return false because it's not exist in the Facebook dataset. This process of FRAppE can be done with 87% of accuracy.
2. In other ways we can check for similarity in the name of apps. If more number of apps seems similar with a malicious app then that apps can be taken as malicious. Otherwise some names can be given as similar but at end of that name they could give the version number that also

can take as malicious apps this is also a valid technique to find the malicious apps with FRAppE. Also we can check for the similarity in the link URL. If the posted link name is similar to the malicious URL, so easily we can identify the malicious apps.

3. At last, we are left with 157 apps that has not identified by the above technique. That apps could verified manually like check one by one and can be identified by using the similarity among this apps and can be identified more than 112 apps which is malicious using FRAppE.

VI.SOCIAL MALWARE ECOSYSTEM

By using FRAppE, we discover the harmful apps , after that we check the several ways how the social malware support each other. From our observation we find the interesting thing that malicious apps do not operate in segregation they share the same name and their work must collaboratively in encouraging each other.

- The emergent's of AppNets

We observed that more than 6,330 malicious apps in our dataset that emerge in collaborative promotion. In that 2.5% are promoters,58.8% are promotes, and the remaining 16.2%play both roles.

- Piggybacking

The app piggybacking is a approach in which hackers are using this. The facebook's API and there post are harmful post by using popular apps. There are several ways that hackers are benefited by this. The hackers make the user to share the harmful post by offering rewards. They crawl the API from Facebook by hacking the users account; they again post the harmful app in the user's wall. By the app in the request to post the harmful post. The Facebook could not recognize this because the app ID is already included in the appID.

In our dataset we identify the piggyback that is each app has atleast one malicious post according to myPageKeeper and we will check for the apps which is having low rates and we found that 80% apps have harmful posts to all posts rate i.e less than 0.4.

VII.CONCLUSION

Here we propose of how safeguard Facebook users from hackers. Using this paper we can understand the significant characteristics of malicious apps and how they operate. In this work we find that atleast 15 % of apps on our dataset are malicious. Malicious apps are differing from benign ones. That is most of the malicious apps have similar name. Profiling each of our observations, we designed FRAppE, a correct classifier for detecting malicious apps on Facebook. To develop FRAppE we use information gathered by observing the posting behavior of 111k Facebook apps seen across 2.2 million users on Facebook. We identify a set of features that help us to distinguish malicious apps from benign ones. And finally we explore the ecosystem of malicious Facebook apps and identify mechanism that these apps use to propagate. We will continue to investigate on hackers platform dig deep into their ecosystem to reduce the malicious app on Facebook.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Mr R Balakrishna(Principal) of Rajarajeswari College of Engineering ,Bengaluru and Dr.Usha Sakthivel (H.O.D) of Rajarajeswari College of Engineering ,Bengaluru. This work was also supported by grants from our Asst.Prof.Mr.V.M.Saravana Perumal (Project Guide) of Rajarajeswari College of Engineering, Bengaluru who helped during various stages of preparations and also provided valuable feedback for guidance

REFERENCES

[1] C. Pring, "*100 social media statistics for 2012*," 2012 [Online]. Available:http://thesocialskinny.com/100-social-media-statistics-for-2012

[2] Facebook, Palo Alto, CA, USA, "*Facebook Opengraph API*," [Online]. Available: http :/ / developers. f a cebook..com/docs/reference/api/

[3] "*Wiki: Facebook platform*," 2014 [Online]. Available: http://en wikipedia.org/wiki/Facebook_Platform

[4] *"Whiich cartoon character are you—Facebook survey scam*," 2012 [Online]. Available: https://appsn.face book.com/ mypagekeeper/?status= sca m _ report_fb_survey_scam_whiich_ cartoon_character_are_you_2012_03_30

[6] E. Protalinski, "*Facebook kills app directory, wants users to search for apps,*" 2011 [Online]. AvaI lable: http://zd.net/MkBY9k

[7] SocialBakers, "SocialBakers: The recipe for socIal ma arketing success," [Online]. Available: http://www.socialbakers.com/

[8] "*Selenium—Web browser automation*," [Online]. Ava ilable:http://seleniumhq.org/

[9] "bit.lyAPI,"2012[Online].Available: http://code.g oogle.com/p/bitlyapi/wiki/ApiDocumentation

[10] Facebook, Palo Alto, CA, USA, "*Permissions r e f e r ence*," [Online].Available: https://developers. facebook.com/docs/authentication/ permissions/

[11] Facebook, Palo Alto, CA, USA, *"Facebook develope rs*," [Online].Available: https://developers.facebook . com/docs/appsonfacebook/tutorial/

[12] "*Web-of-Trust,"* [Online]. Available: http://www.myw
t.com/

[13] F. J. Damerau, "*A technique for computer detection and correction of spelling errors*," *Commun. ACM*, v
ol. 7, no. 3, pp. 171–176,Mar. 1964.

[14] C.-C. Chang and C.-J. Lin, "*LIBSVM: A library for su
pport vector machines," Trans. Intell. Syst. Technol.*,
vol. 2, no. 3, 2011, Art. no. 27.

[15] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "*B
eyond blacklists: Learning to detect malicious Web si
tes from suspicious URLs,"* in *Proc. KDD*, 2009, pp.
1245–1254.

[16] A. Le, A.Markopoulou, and M. Faloutsos, "*PhishDef:
URL names say it all*," in *Proc. IEEE INFO COM*, 2011, pp. 191–195.

[17] C. Wueest, "*Fast-flux Facebook application scams*,"
2014 [Online].Available:http://www.symantec.com/
connect/blogs/fast-fluxfacebook- application-scams

[18] "*Longest path problem*," 2014 [Online]. Available: htt
p://en.wikipedia. org/wiki/Longest_path_problem

[19] "*App piggybacking example*," [Online]. Available: htt

ps://apps:Facebook.Com/mypagekeeper/?status=sca
m _report_fb_survey_scam_ Converse_shoes_2012
_05_17_boQ

[20] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song,
*Design and evaluation of a real-time URL spam filt
ering service*," in *Proc. IEEE Symp. Security Privacy*, 2011, pp. 447–462.

[21] S. Lee and J. Kim, "*WarningBird: Detecting suspicious
URLs in Twitter stream*," in *Proc. NDSS*, 2012.

[22] C. Yang, R. Harkreader, and G. Gu, "*Die free or live h
ard? Empirical evaluation and new design for fighting evolving Twitter spammers*," in *Proc. RAID*, 2011, pp. 318–337.

[23] F. Benevenuto, G. Magno, T. Rodrigues, and V. A l m
ida, "*Detectin spammers on Twitter*," in *Proc. CEAS*,
2010, pp. 1–9.

[24] G. Stringhini, C. Kruegel, and G. Vigna, "*Detecting sp
ammers on social networks*," in *Proc. ACSAC*, 2010, p
p.1–9.

[25] K. Lee, J. Caverlee, and S.Webb, "*Uncovering social
spammers: Social honeypots + machine learning*," in
Proc. SIGIR, 2010, pp. 435–442.

[26] S.Yardi, D.Romero, G.Schoenebeck,and D.Boyd, "*Detecting spamin a twitter network*,"*First Monday*, vol.
15,no.1,2010[Online]

[27] A. Besmer, H. R. Lipford, M. Shehab, and G. Cheek, "
Social Applications;Explorin a more secure frame work," in *Proc. SOUPS*, 2009,Art. no. 2.

[28] N. Wang, H. Xu, and J. Grossklags, "*Third-party apps
on Facebook:Privacy and the illusion of control*
," in *Proc. CHIMIT*, 2011, Art. no.4

[29] A. Makridakis *et al.*, "*Understanding the behavior of m
alicious Applications in social networks," IEEE Ne tw.*, vol. 24, no. 5, pp. 14–19, Sep.–Oct. 2010.

[30] J. King, A. Lampinen, and A. Smolen, *"Privacy: Is th
ere an app for that?,"* in *Proc. SOUPS*,
2011, Art. no. 12.

[31] M. Gjoka, M. Sirivianos, A. Markopoulou, and X.
yang , "Poking *Facebook: Characterization of OSN
applications*," in *Proc. 1st WOSN*, 2008, pp. 31–36.

[32] T. Stein, E. Chen, and K. Mangla, "*Facebook immune
system*," in *Proc. 4th Workshop Social Netw. Syst.*, 2011, Art. no. 8.

[33] L. Parfeni, *"Facebook softens its app spam controls, In
troduces better tools for developers*," 2011
[Online]. Available: http://bit.ly/LLmZpM

[34] *"Norton Safe Web*," [Online]. Available: http ://www
.facebook.com/ apps/application.php?id=
310877173418

A Novel Approach for Efficient Data Collection in Cluster Based Communication using Compressive Sensing and MIMO Techniques to Extend Network Lifetime

Sunanda Patne, Vinodha K
Department of Information Science, Visvesvaraya University, Karnataka-, 18
sunandapatne2009@gmail.com, mohan.vinodha@gmail.com

ABSTRACT- In Wireless Sensor Networks (WSNs) consumption of energy is a major concern. In this paper Cluster Based Communication using Compressive Sensing and Multi Input Multi Output (MIMO) Techniques for efficient data collection is proposed. This system consists of a three level framework which includes the cluster head layer, sencar (mobile collector) layer, and sensor layer. By using this framework good scalability, prolong network lifetime and low data collection delay is achieved. At cluster head layer, a distributed load balanced clustering algorithm is proposed. Our scheme generates multiple cluster head in each cluster to balance the load and provides dual data uploading. At the cluster head layer, the connectivity among the clusters is guaranteed by choosing appropriate inter-cluster transmission range. To make the system energy efficient multiple cluster head within the cluster they cooperatively communicate with each other. Before uploading, the data is compressed by the Cluster Head using compressive sensing technique to reduce data transmission time. By using compressed data sensing and dual data uploading technology the proposed system achieves over 50 percent energy saving per node and 60 percent energy saving on cluster heads comparing with data collection through multi-hop relay to the static data sink, and 20 percent shorter data collection time compared to traditional mobile data gathering.

Keywords – Mobile Collector, Polling Point, Multi Input Multi Output, Relay Node, Dual Uploading, Load Balancing.

1. Introduction

Wireless Sensor Networks (WSNs) have emerged as a new information-gathering paradigm in a wide range of applications, such as medical treatment, outer-space investigation, battlefield surveillance, emergency response, etc. where Sensor nodes are usually thrown into a large-scale sensing field without a preconfigured infrastructure. Before monitoring the environment, sensor nodes must be able to discover nearby nodes and organize themselves into a network. The measuring and monitoring of the data packets from multiple sensors are then processed and forwarded to external networks via static or mobile sink which act as gateways. This procedure is called data gathering. Sensors are used to sense, process and record conditions in different locations. Every sensor node has a power source typically in the form of a battery. The base stations are one or more components of the WSN with infinite energy and communication resources. They act as an interface between sensor nodes and the end user as they typically forward data from the WSN to a server. Clustering is particularly useful for applications with scalability

requirement and is very effective in local data aggregation, since it can reduce the collision and balances the load among sensors. The sencar (mobile collector) collects the aggregated data from the cluster heads by visiting predefined locations named as polling points through a planned trajectory.

This paper is organized as follows. Section 2 deals with the system overview, related work is presented in section 3. The detailed implementation methodology of the proposed system is discussed in section 4; the performance evaluation of the system is discussed in section 5. The result analysis is done in section 6, finally the concluding remarks and future work is presented in section 7 and 8 respectively.

2. System Overview

An overview of cluster based communication of three layer framework is shown in Fig.1, which consists of three layers: sensor layer, cluster head layer and sencar layer. The sensor layer is the bottom and basic layer. In the Initial set up phase the sensors self organize themselves into a cluster. Each individual sensor decides by itself to be a cluster head or member of the cluster. The sensor with high residual energy becomes the cluster head and each cluster has at most M cluster heads, where M is the positive integer [1]. The cluster heads act as peers to each other. The algorithm places these sensors in such a way that they are one hop away from at least one cluster head. The benefit of such organization is that the intra-cluster aggregation is limited to a single hop. To avoid collisions during data aggregation, the cluster head group (CHG) adopts time-division-multiple-access (TDMA) based technique to coordinate communications between sensor nodes. After the election of cluster heads, the nodes synchronize their local clocks via beacon messages. For example, all the nodes in a CHG adjust their local clocks based on the node with the highest residual energy.

The cluster head layer consists of all the cluster heads and member nodes. The inter-cluster forwarding is only used to send the CHG information of each cluster to sencar, which contains a list of ids of multiple cluster heads as well as the priority of the data to be sensed in a CHG. Such information must be sent before sencar departs for its data collection tour [1].Upon receiving this information; it plans the path trajectory for data collection from the CHs. To guarantee the connectivity for inter-

cluster communication, the cluster heads in a CHG can cooperatively send out duplicated information to achieve spatial diversity, which provides reliable transmissions and energy saving. The top layer is the sencar layer, which mainly manages mobility of sencar. There are two issues to be addressed at this layer. First, we need to determine the positions where sencar would stop to communicate with cluster heads when it arrives at a cluster. In the proposed algorithm, sencar communicates with cluster heads via single-hop transmissions. It is equipped with two antennas while each sensor has a single antenna. The traffic pattern of data uploading in a cluster is many-to-one, where data from multiple cluster heads converge to sencar. It is equipped with two receiving antennas, for dual data uploading whenever possible, in which two cluster heads can upload data simultaneously. By processing the received signals with filters based on channel state information, sencar can successfully separate and decode the information from distinct cluster heads. To reduce the latency in collecting the data the sencar stops at predefined polling point such that it encounters maximum number of CHs that comes under its transmission range. Since sencar is mobile, it has the freedom to choose any preferred position. However, this is infeasible in practice, because it is very hard to estimate channel conditions for all possible positions. Thus, we only consider a finite set of locations. To mitigate the impact from dynamic channel conditions, sencar measures channel state information before each data collection tour to select candidate locations for data collection. We call these possible locations as polling points. Since sencar has pre-knowledge about the locations of polling points, it can find a good trajectory by seeking the shortest route to visit each selected polling point exactly once and then returns to the data sink.

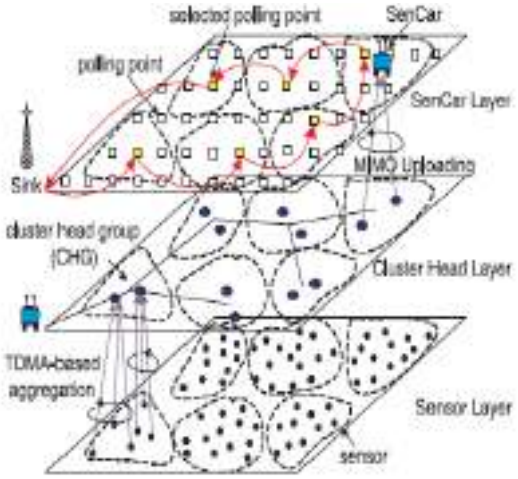


Fig. 1 Illustration of three layer framework

3. Related Work

Clustering Scheme such as a relay routing, is an effective and simple way to approach the concept of routing messages to the data sink in multi-hop fashion. Cheng et al[3] discovered the alternate route method to avoid congestions and transfer data. Wu et al. [4] studied to maximize the life time of data storage by designing an algorithm that started with a random design and ideally ended with load reduction on the bottleneck nodes. Xu et al. studied on the relay approach towards nodes to extend the lifetime of the network. The cluster tree protocol (CTP) is mainly used to work on robustness, reliability, efficiency and hardware independent. Whatever maybe, but when it comes to nodes in the critical data path, if energy depletion occurs then data collection performance will decline. Heinzelman et al. [5] provided another applicable approach using the concept of clusters called LEACH. LEACH leads to less number of relays and instead forms the cluster groups. However the performance of the cluster depends on the cluster head. Younis and Fahmy [4] proposed the concept of HEED, which helps in the cluster head selection. The combination of residual energy and cost is considered in this approach. HEED can deliver well-distributed cluster heads and compact clusters. Gong et al [6] discovered energy efficient clustering in wireless sensor networks through quality links. Amis et al [7] contributed towards another interesting concept of d-hop cluster where each node is at d hop away from the cluster head. Among all the cluster based models, clusters not only act as a local data collector, a cluster head also acts as a controller and a scheduler for in-network process. Cluster heads use the spatio-temporal concept to minimize the reading process to maximize the energy saving. Single head clustering schemes are not compatible with Multiple Uploading-Multi Input Multi Output (MU-MIMO).In this paper a load balanced; multi-head clustering algorithm is proposed to improve data collection latency and network lifetime.

4. Proposed Methodology

In the applications of WSN, the sensors are densely deployed and scattered over a sensing field and left unattended once being deployed, that makes it tough to recharge or replace their batteries. Once sensors starts working, the sensors which are close to the information sink usually deplete their battery power at a faster rate compared to the other sensor node, due to additional relaying traffic. Once sensors round the information sink [2].deplete their energy, network property and coverage might not be secured. The main aim is to propose an energy efficient system that consumes energy uniformly across the sensing field to attain long network time period. Moreover, as sensing information in some applications are time-sensitive, information collection is also needed to be performed inside a fixed time-frame. Therefore, a large-scale information collection scheme aims at a smart, long network time period and low information latency scheme. Many approaches were discussed for economical information collection within the literature, in this work we discussed a Cluster Based communication using compressive sensing and MIMO techniques to extend network lifetime.

4.1. Sensor Layer – Load Balanced Clustering:

In sensor layer, the sensors self-organize themselves into clusters. The essential operation of clustering is the selection of cluster heads. The selected cluster heads are the ones with higher residual energy. Each sensor is covered by at least one cluster head inside a cluster. Clustering enables network scalability and extends the life of the network by allowing the sensors to conserve energy through communication with closer nodes and by balancing the load among the cluster head nodes. Clusters are formed based on the cost of communication and the load on the cluster heads.

The Load balanced clustering (LBC) algorithm is comprised of four phases

· Initialization

· Status claim

· Cluster forming

· Cluster head synchronization

4.1.1. Initialization:

The Initialization is done at the sensor layer using LBC algorithm. The sensor informed its existence to all neighbors within its transmission range. If a sensor has no neighbor existence, it claims itself to be cluster. Otherwise sensor sets its status as tentative and its priority set by the percentage of residual energy. Then it sorts the neighbors with high residual energy as candidate peers. In the initialization phase, each sensor acquaints itself with all its near neighbors. The sensors $_{Si}$ would pick one neighbor with the highest initial priority as its candidate peer [8].

Algorithm 1. Phase I: Initialization
1: My. N ← {v | v lies in my transmission range; v € S;
2: if My. N Φ then
3: Set My. cluster_head to My.id;
4: Set My. Status to cluster_head;
5: else
6: My. init_prio ← Eres=Etot;
7: My. cluster_head ← 0;
8: My. Status ← tentative;
9: My. A ← v| v € Can_Peers (N)};
10: My. prio ← My.init_prio+∑v € My: A v:init_prio;
11: My. B, My.C _Φ;
12: Iter _ 0;

4.1.2. Status Claim Phase:

Each sensor determines its status by iteratively updating its local information, refraining from promptly claiming to be a cluster head. The sensors are members in the cluster. We use the node degree to control the maximum number of iterations for each sensor. The priority is partitioned into two thresholds τh(threshold of CH), τm (threshold of cluster member node) this is used to declare a sensor as either cluster head or cluster member [8].

4.1.3 Cluster Forming Phase:

The third phase is cluster forming that decides which cluster head a sensor should be associated with cluster consists of two cluster heads and sensors. The sensor with tentative status, or being a cluster member, is arbitrarily chosen as the cluster head from its candidate peers for load balancing purpose. If no sensor with tentative status then it chooses itself as the cluster head. The re-clustering is performed when the chosen cluster head is running on low battery. The Initialization phase is done by sending re-clustering messages to all sensors. The following algorithm explains about how clustering is done and how they receive packets from the other sensor.

Let us consider two cluster head i.e. C1, C2 whose cluster members are m1, m2, m3 and m4 respectively and τh be the threshold value. Let us assume that the initial energy of CH as xJ. Assign each cluster with energy of E and the maximum energy each cluster head is say $E_{max.}$

if $E(C1) > E_{max}$ (m1, m2)

and E τh(C1) > E_{max} (m1, m2)

then C1 is the cluster head.

Similarly, if E (C2) > E_{max} (m3, m4)

And E τh(C2) > E_{max} (m1, m2)

Then C2 is the cluster head.

Therefore C1 and C2 are declared as cluster head and m1, m2, m3 and m4 are the members of the cluster head.

The Cluster Formation includes the following steps:-

1. Number of clusters, vertical and horizontal end points for each of the cluster acts as an input.
2. For each of the cluster formation Step3, 4 and 5 are repeated until all clusters are formed
3. Pick appropriate x end points and y end points for the cluster.
4. Generate the cluster id for the cluster.
5. Execute the node deployment algorithm and place the nodes within the cluster.

4.1.4. Cluster Head Synchronization Phase:

To perform data collection by TDMA techniques, intra-cluster time synchronization among established cluster heads should be considered. The fourth phase is to synchronize local clocks among cluster heads in a CHG by beacon messages [1]. The communication between the nodes and cluster head in the CHG is called intra cluster communication. The synchronization among cluster head is done because to perform data collection by time division. This is done by sending beacon messages to cluster heads in CHG. the message contains the local clock information and initial priority. This is done only when sencar is going to collect data.

Algorithm: Synchronization between cluster heads

Step 1: if (status=cluster_head) then

Step2: Send initial priority, clock messages, etc

Step3: Receive beacon msg b, from other cluster heads.

Step 4: Compare b msg with sent msg

Step 5: if b msg is > status msg

Step 6: Set/Adjust the beacon msg clock to my clock.

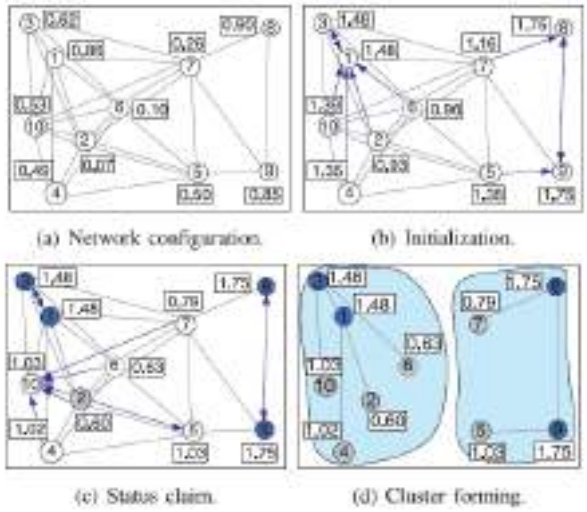


Fig. 2 An algorithm of load balanced clustering with m=2

MU-MIMO in WSN:

The feasibility of employing MIMO techniques in wireless sensor networks is envisioned [9][10][11]. Due to difficulties to mount multiple antennas on a single sensor node, MIMO is adopted in WSN to seek cooperation from multiple nodes to achieve diversity and reduce bit error rate. MIMO-based scheduling algorithm is used to coordinate transmissions. MU-MIMO can greatly speed up data collection time and reduce the overall latency. Another application scenario is in disaster rescue. For example, to combat forest fire, sensor nodes are usually deployed densely to monitor the situation. These applications usually involve hundreds of readings in a short period (a large amount of data) and are risky for human being to manually collect sensed data. A mobile collector equipped with multiple antennas overcomes these difficulties by reducing data collection latency and reaching hazard regions not accessible by human being. Although employing mobility may elongate the moving time, data collection time would become dominant or at least comparable to moving time for many high-rate or densely deployed sensing applications. In addition, using the mobile data collector can successfully obtain data even from disconnected regions and guarantee that all of the generated data is collected.

5. Implementation Methodology

Network Simulator-2 is used for the simulation, to implement the probing mechanism, congestion detection, and node placement algorithms. We create a sensor network of 50 nodes that use a single channel for communication. The nodes are configured to use the 802.11C protocol with a maximum data rate of 25Mbps and control rate of 10Mbps and a transmission range of 500mts. Based on the priority of the data sensed the sencar collects the data from the cluster heads by moving through the planned trajectory.

The implementation phase includes three modules, they are as following:

MODULES

1. Node Construction
2. Cluster Head Formation
3. Data Transmission through sencar

5.1. Node Construction:

The NS2 simulator is used to simulate the proposed algorithm. The number of nodes considered is 50 nodes, in which each node moves around 50m/s, with transmission range of 100m, bit rate of 250bits/s, with initial energy level of 100J and then sensors are positioned. The parameters used for simulation is as listed in Table 1.

Table 1.Parameters used for simulation:

PARAMETERS	DESCRIPTION
CHANNEL	WIRELESS
RADIO PROPAGATION	TWO-RAY GROUND
DEPLOYMENT AREA	300×300m to 1000m×1000m
ANTENNA TYPE	OMNI TYPE
MAC LAYER	802.11.C
NO. OF NODES	50,100 etc
THRESHOLD	250bits/s
TOPOLOGY	FLAT GRID
TRANSMISSION RANGE	50m
ENERGY	100 Joules

5.2. Cluster Head Formation:

In this module, each sensor determines its status and the sensor which has high residual energy become a cluster head and other sensors are members in the cluster. Sensors are self organize into the cluster by the LBC algorithm. Each cluster consists of two cluster heads and

sensors. The residual energy of each and every node is broadcasted to the neighbors, so which ever node is having the higher residual energy is considered as a CH. The cluster members who are at one hop distance from cluster head they become the member of that CH. The CH once it is been elected it sends it id to the sencar and also based on the priority is been sensed.

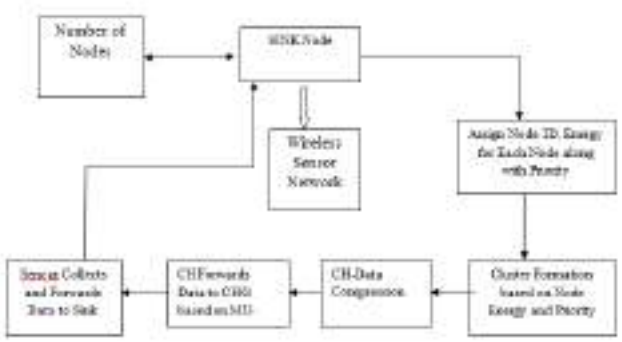


Fig 3. Proposed system architecture

5.3. Data transmission through sencar:

Source node in cluster head sends data to base station via group cluster head and sencar node. As in Fig 3, in that process, sensor nodes send data to its cluster head. Then the cluster head sends the collection of data to its group cluster head. We coordinate the mobility of sencar to fully enjoy the benefits of dual data uploading, which ultimately leads to a data collection tour with both short moving trajectory and short data uploading time. Finally sencar node collects the data from group cluster head and gives that collected data to base station.

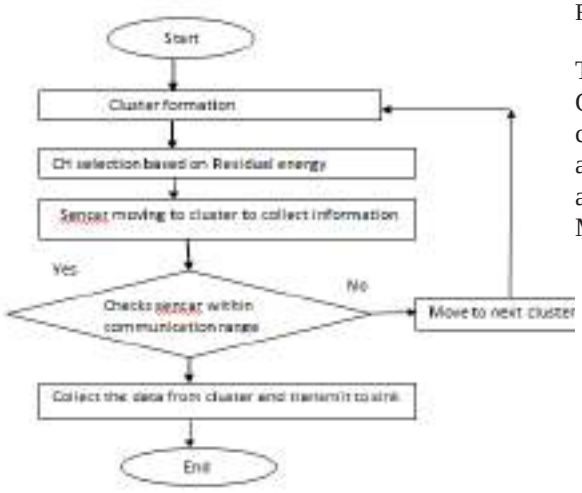


Fig 4. Flow chart for data collection by sencar

A data flow diagram (DFD) is a logical model of the flow of data through a system that shows how the system's boundaries, processes, and data entities are logically related. In above fig 4, first step is formation of the cluster(i.e. group of nodes),then we select the cluster head based on the residual energy(left out energy), the mobile collector keeps on moving for collecting the information from the cluster heads, which are within the transmission range of SenCar.

6. Performance Evaluation and Result Analysis

In order to evaluate the performance of proposed algorithm, we performed a simulation in NS2 and collected the data. The metrics used for the comparison of existing clustered SISO with the proposed MU-MIMO are delay, over head and energy consumption. The MIMO scheme results in least energy consumption so the lifetime of the network also extended, because the sensor sends the data transmission by multi hop fashion. The low latency in data uploading is achieved by using dual data uploading technique in sencar.

The fig 5, Shows the energy consumption of MU-MIMO with compressive sensing and without compressive sensing in WSN. The data sensed by the member sensor nodes may contain redundant data. So the CHs compress the received data by using the redundancy compressive technique and upload the data to sink. This results in fast uploading of data which deals with results in less delay and less energy consumption.

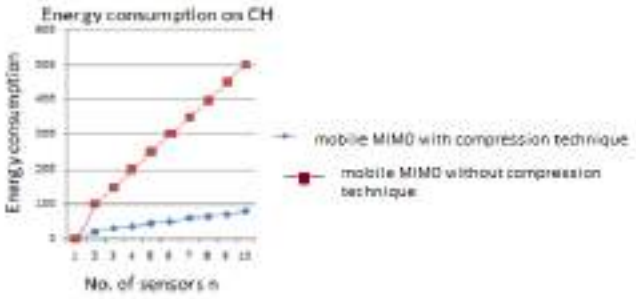


Fig. 5 Compression technique in mobile MIMO

The Fig 6. explains about the energy consumption on each CH.In clustered Single Input Single Output(SISO) each cluster has single CH, but in MU-MIMO, each cluster has atmost two CH.So the load is balanced between two CH and therefore energy consumption is reduced in MU-MIMO.

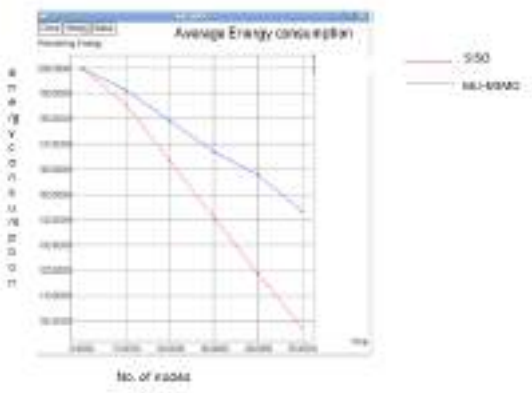


Fig.6 Energy consumption in SISO and MU-MIMO

The Fig. 7 explains about the delay. In clustered SISO the mobile collector is equipetted with one antenna hence it can receive data from one CH at a time. This increases the

delay.But in MIMO, equippied with two antenna, for two CH so it can receive the data simulataneously from two CH. In order to avoid interference the uploading of data is based on time schedule

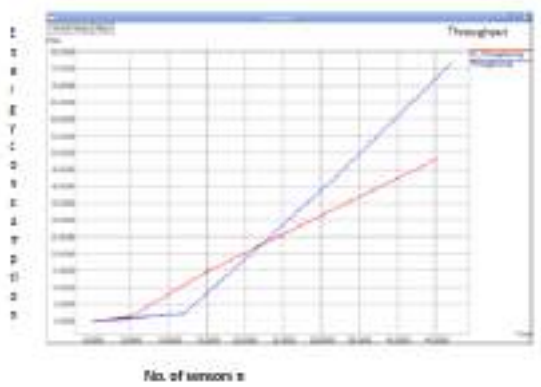


Fig.7 Delay in SISO and MU-MIMO

7. Conclusion

The cluster based communication using compressive sensing and MIMO techniques is proposed in this paper. It consists of sensor layer, cluster head layer and sencar layer. It employs distributed load balanced clustering for sensor self-organization, adopts collaborative inter-cluster communication for energy-efficient. Transmission among cluster Head Groups, uses, dual data uploading for fast data collection, and optimizes SenCar's mobility to fully enjoy the benefits of MU-MIMO. Our performance study demonstrates the effectiveness of the proposed framework in terms of packet delivery ratio, delay and energy consumption.

8. Future Scope

There are some interesting problems that may be studied in our future work. The first problem is how to find polling points and compatible pairs for each cluster. A discretization scheme should be developed to partition the continuous space to locate the optimal polling point for each cluster. Then finding the compatible pairs becomes a matching problem to achieve optimal overall spatial diversity. The second problem is how to schedule MIMO uploading from multiple clusters. An algorithm that adapts to the current MIMO-based transmission scheduling algorithms should be studied in future.

REFERENCES

[1] Mobile Data Gathering with Load Balanced Clustering and Dual data gathering with load balanced clustering in wireless sensor network,Miao Zhao, member, Yuanyuan Yang, Fellow,IEEE and Yang, Fellow,IEEE and Cong Wang, *IEEE transaction on mobile computing,Vol. 14,no. 4,april 2015*

[2] A load balancing in clustering with data gathering for network prolong using sencar in wireless sensor network *International Journal for Technological Research In Engineering Volume 3,* Issue 4, December-2015

[3] W. C. Cheng, C. Chou, L. Golubchik, S. Khuller, and Y. C. Wan, "A coordinated data collection approach: Design, evaluation, and comparison," *IEEE J. Sel. Areas Commun., vol. 22, no. 10, pp. 2004–2018, Dec. 2004.*

[4] Y. Wu, Z. Mao, S. Fahmy, and N. Shroff, "Constructing maximum- lifetime data-gathering forests in sensor networks," *IEEE/ ACM Trans. Netw., vol. 18, no. 5, pp. 1571–1584, Oct. 2010.*

[5] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun., vol. 1, no. 4, pp. 660– 660, Oct. 2002.*

[6] D. Gong, Y. Yang, and Z. Pan, "Energy-efficient clustering in lossy wireless sensor networks," *J. Parallel Distrib. Comput., vol. 73, no. 9, pp. 1323–1336, Sep. 2013.*

[7] A. Amis, R. Prakash, D. Huynh, and T. Vuong, "Max-min d-cluster formation in wireless ad hoc networks," in *Proc. IEEE Conf. Comput. Commun., Mar. 2000, pp. 32–41.*

[8] A load balancing in clustering with data gathering for energy efficient using sencar in wireless sensor networks. *International Conference on Explorations and Innovations in Engineering & Technology (ICEIET – 2016)*.

[9] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," IEEE J.Sel. Areas Commun., vol. 22, no. 6, pp. 1089–1098, Aug. 2004

[10] S. Jayaweera, "Virtual MIMO-based cooperative communication for energy-constrained wireless sensor networks," *IEEE Trans. Wireless Commun., vol. 5, no. 5, pp. 984–989, May 2006.*

[11] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun., vol. 4, no. 5,pp. 2349–2360, Sep. 2005.*

[12] X. Tang and J. Xu, "Adaptive data collection strategies for lifetime- constrained wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst., vol. 19, no. 6, pp. 721–7314, Jun. 2008.*

.**BIOGRAPHY:**

Ms. Sunanda Patne received her Bachelor of Engineering in Computer Science and Engineering in 2010. Currently she is a M.Tech student in Computer Networking Engineering from Visvesvaraya Technological University at The Oxford Collage of Engineering, Bangalore. Her research interests are Internet of Things (IOT), Computer Networks, and Network Security.

Mrs. Vinodha K received her Bachelor of Engineering in Electronics and Communication in 1998. She received her ME in Computer Science and Engineering from Anna University in 2008.She is pursuing Ph.D in Computer Science and Engineering at VTU, Belgaum. Currently she also holds a faculty position as Assistant Professor, Department of ISE, The Oxford College of Engineering. Her main research interests are Computer Networks and Distributed Networks

Management of QoS in Sensor Cloud

Anasuya N J[1], Akhila S J[2]
Associate Professor, PG Scholar, Department of CSE, DBIT Bangalore
anuprama@rediffmail.com, sjakhila.05@gmail.com

Abstract—Sensor cloud is a new concept in cloud computing which provides sensor management platform that works as interface between physical and internet world. It is a collection of sensor nodes which are activated on demand and forms a group virtual sensors (VS). These VSs are located in different locations which transfers collected data into nearest data center. The Proposed System concentrates on Scheduling a particular DC which assembles data from various VS and serves the request of clients/end users. Scheduling a particular DC depends on several network constraints such as Average processing time, data delivery cost, service delay of an application and QoS. Simulation results areshown for the proposed system.

Keywords—WSN, Cloud Computing, Sensor Cloud, Data Center

I INTRODUCTION

Recent research has acknowledged the sensor-cloud infrastructure as a potential substitute of traditional Wireless Sensor Networks (WSNs) [2][3]. Now day's sensors are used for various fields such as environment, healthcare and government services. Whenever user requests for sensor information, Cloud computing IT resources can provide users with virtual servers [3]. By using this virtual sensors users need not worries about the locations of the servers. Sensor Cloud infrastructure uses numerous physical sensors to form virtual sensor. Sensor cloud is a new concept in cloud computing [2], which is known as a sensor management platform that acts as an intermediate layer between physical and internet world (cyber world). Sensor Cloud infrastructure uses physical resources (i. e. sensors) within the cloud environment and delivers Sensors-as-a Service (SeaaS) to end-users .Hence such a new technology permits the end users to visualize the sensor nodes as a service, rather than hardware. Sensor services are divided into sensor system management and sensor data management [2]. Sensor Cloud infrastructure explains sensor system management. The existing system focuses on accepting one or more clients request for various types of sensor data which are located in different regions in the form of sensor as a service (SeaaS). For every request from end user, distinct VS are formed, the data from the various VS are transfer into nearest DC for temporary storage and all data from temporary DC are transmitted into a randomly selected DC which can be used to serve for end user [1]. This random selection of DC leads to reduction in QOS, more traffic delay and chances of loss of data. So goal of the work is to design a system which selects a particular DC that serves a user application

II RELEATED WORK

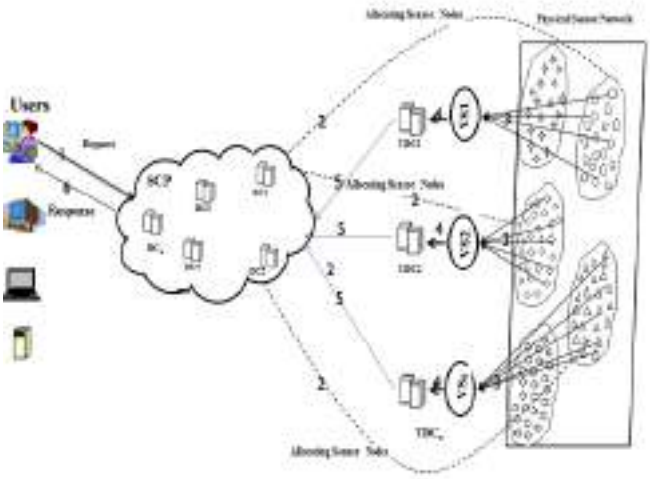
Prior to this work, [1] address the problem of scheduling a DC which collects the information from several VSs and transfers the same to the user who sends request. The work is based on general pairwise choice framework. Sensor-Cloud Infrastructure achieves physical sensor on IT infrastructure [2][3]. This Sensor-Cloud infrastructure consists of virtual sensors which send sensed information to cloud. Virtual Sensor groups can be formed as and when user needs them [10] [4]. A framework of sensor-cloud connection to use the sensor data and also a content-based pub-sub model which shortens the incorporation of sensor network with cloud based community and emphases on sensor cloud theoretical modeling [5] [6]. In addition to this [7] addresses the problem that how to track multiple targets in Sensor-Cloud. A highly efficient tabu search algorithm is proposed for optimizing locations of cloud datacenters, software components and a planning problem [8] and also a novel approach which based on a Markov chain model that solves the issue of detection of overloaded host which depends on calculating the mean inter migration time under given QoS value and they also uses a Multisize Sliding Window workload estimation technique to handle unknown non-stationary workloads [9]. R. C. Ben-Yashar and S. I. Nitzan describes how to calculate the optimal decision rule based on general pairwise choice framework. And this works assumes four assumptions while calculating optimal decision rule.

III PROBLEM SCENARIO

The existing system focuses on accepting one or more clients request for various types of sensor data which are located in different regions in the form of sensor as a service (SeaaS). For every request from end user, distinct VS are formed.

The data from the various VS are transfer into nearest DC for temporary storage and all data from temporary DC are transmitted into a randomly selected DC which can be used to serve for end user [1]. This can be shown in fig 1. The above figure explains

1. Sending request to sensor cloud
2. Sensor cloud provider allocating the physical sensor as per the user request
3. Formation of virtual sensor (VS) group
4. Transferring data from VS to temporary data centers
5. Migrating sensed data from temporary data centers to permanent data centers present in sensor cloud
6. Sending responses to end user.



SCP- Sensor cloud provider
TDC- Temporary Data Center
DC- Data Center, VS- Virtual Sensor

Fig.1 Working of Existing System

PROPOSED SYSTEM

We are scheduling a DC which is calculated based on several networking constraints such as Migration cost, delivery cost, service delay, QoS.

A. Symbols and Formulae[1]

The formulae used in calculation are as follows

Table1: List of symbols

Symbols	Description
P	Number of packets
P	Size of Packets
d(u,dci)	Distance between user to permanent DC
η_1	Migration Constant
η_2	Transmission rate from DC to an end-user

The average processing time for processing a 'P' number of packets each of size 'p' bytes is given by

Average Processing Time = $(P*p)/\eta_1$ **(1)**

The delivery cost for transferring service response to user is given by

Delivery Cost= $d(u,dci)^2/\eta 2$ **(2)**

Service Delay in transferring service response is

Service Delay = $\sum$ (Average Processing Time + Delivery Cost) **(3)**

Finally QoS can be calculated as

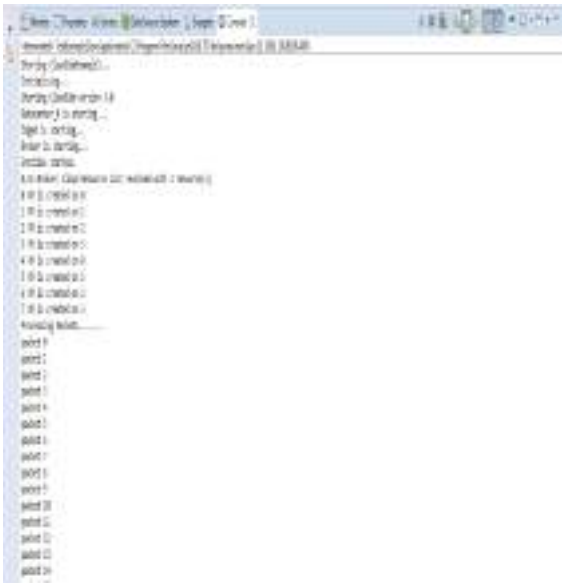
$$\textbf{QoS} = \frac{P^p}{\text{Service Delay}} \quad \textbf{(4)}$$

IMPLEMENTATION

The Cloud Sim 3.0.2 Simulator has Choose to perform the implementation process since it is a modern simulation tool. We have simulated for different number of hosts and VMs, each host contains number of VMs capable of processing a packet. The simulation result is shown in Figure 2.

SIMULATION RESULT

The simulation result is shown in Figure 3. Initially we have simulated for four hosts, eight VMs and hundred tasks. Andparallely calculated Average Processing time, delivery cost, service delay and QoS.



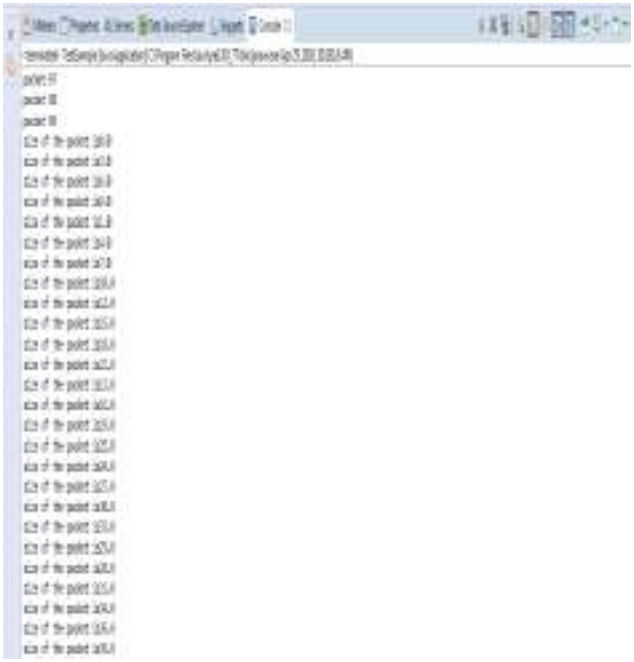


Fig. 2 Showing requests

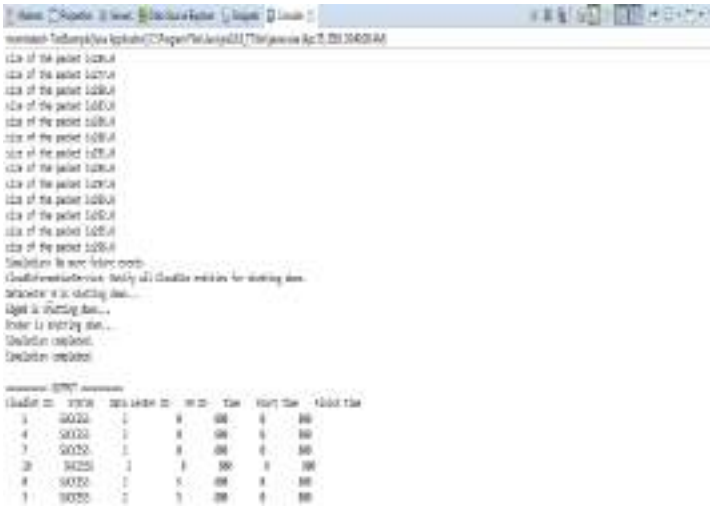


Fig. 3 Displaying size of packets

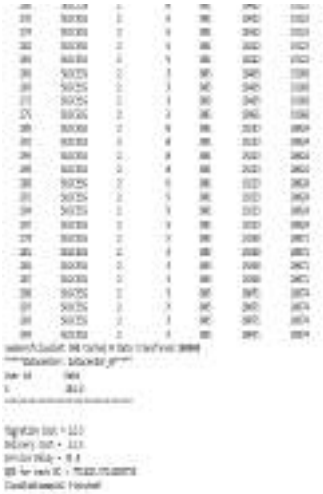


Fig. 4displaying cloudlets

Fig. 5 Displaying Calculated values

CONCLUSION

The proposed work concentrates on sensor cloud infrastructure. In sensor cloud, data /information from all the physical sensors together forms a group called Virtual Sensor (VS) and data from these VSs are transferred into geographically located geospatial DC for temporary storage. Data from this temporary storage are migrated into a single VM present within DC for further processing. This can be done based on scheduling of DC by maximizing QOS for each application/user.

REFERENCES

[1] Subarna Chatterjee, Sudip Misra and Samee U. Khan, "Optimal Data Center Scheduling for Quality of Service Management in Sensor-cloud", in IEEE Transactions in Cloud Computing, 2015

[2] M. Yuriyama and T. Kushida, "Sensor-cloud infrastructure physical sensor management with virtualized sensors onvcloudcomputing," invInternational Conference on Network-Based Information Systemsv, 2010.

[3] M. Yuriyama, T. Kushida, and M. Itakura, "A new model of accelerating service innovation with sensor-cloud infrastructure," in Annual SRII Global Conference (SRII) , March 2011, pp. 308Ì.

[4] S. Madria, V. Kumar, and R. Dalvi, "Sensor cloud: A cloud of virtual sensors," Software, IEEE , vol. 31, no. 2, pp. 70, Mar 2014.

[5] M. M. Hassan, B. Song, and E.-N.Huh, "A framework of sensor-cloud integration opportunities and challenges," in Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, ser. ICUIMC '09. New York, NY, USA: ACM, 2009, pp. 618–. [Online]. Available: http://doi.acm.org/10.1145/1516241.1516350

[6] S. Misra, S. Chatterjee, and M. S. Obaidat, "On theoretical modeling of sensor-cloud: A paradigm shift from wireless sensor network," IEEE Systems Journal (Accepted) , 2014.

[7] S. Chatterjee and S. Misra, "Target tracking using sensor-cloud: Sensor- target mapping in presence of overlapping coverage," IEEE Communications Letters , vol. PP, no. 99, pp. 1±, 2014.

[8] F. Larumbe and B. Sanso, "A tabu search algorithm for the location of data centers and software components in green cloud computing networks," IEEE Transactions on Cloud Computing , vol. 1, no. 1, pp. 22•, Jan 2013.

[9] A. Beloglazov and R. Buyya, "Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints," IEEE Transactions on Parallel and Distributed Systems , vol. 24, no. 7, pp. 1366_, July 2013.

[10] S. Misra, A. Singh, S. Chatterjee, and M. Obaidat, "Mils-cloud: A sensor-cloud-based architecture for the integration of military tri-services operations and decision making," IEEE Systems Journal , vol. PP, no. 99, pp. 1¹, 2014.

[11] R. C. Ben-Yashar and S. I. Nitzan, "The optimal decision rule for fixed-size committees in dichotomous choice situations: The general result," International Economic Review , vol. 38, no. 1, pp. pp. 175†, 1997. [Online]. Available: http://www.jstor.org/stable/2527413

Controlled Data Access Using IBE in Cloud Computing

Bhoomika M U, Chandini M, Harshitha.R.S, Suraj S Gurav, Prof. Mangala.C.N
UG SCHALOR, Associate Professor
Department of CS&E,
East West Institute of Technology, Bangalore, India
bhoomi123mu@gmail.com, chans.1994@gmail.com, harshithasiddesh93@gmail.com, guravsuraj@gmail.com
mangalacn@ewit.edu

***Abstract* –** Identity Based Encryption or Identity based encryption is an important primitive of ID-Based Cryptography. It is an important alternative to public key encryption. Identity based encryption which simplifies the public key and certificate management. Propose a revocable Identity based encryption scheme in the server aided setting. It is achieved by utilizing a novel collusion resistant technique which means generating a hybrid private key for every user using AND gate it helps to connect and bound the identity and time component. In this paper aiming at tackling the critical issue of identity revocation we introduced the outsourcing revocation for the first time and the purpose of revocable IBE scheme in server aided sitting. Our scheme offloads most of the generation related operations during key issuing and key update process to a key update cloud service provider, leaving only a constant number of simple operations for PKG and user to perform locally. This goal is achieved by utilizing a novel collusion resistant technique: we employ a hybrid private key for each user, in which an AND gate is involved to connect and bound the identity component and the time component. Furthermore it gives a secure under the recently formulised refereed delegation of computation model. Finally we provide extensive experimental results to demonstrate the efficiency of our proposed construction.

Index terms-Identity based encryption, key generation, cloud computing.

1. INTRODUTION

Identity based encryption is an important primitive of ID-based cryptography. As such it is a type of public key encryption in which the public key of a user is some unique information about the identity of the users (e.g. a user's email address).this means that a sender who has to the public parameters of the system can encrypt a message using e.g. the text-value of receiver's name or email address as a key .the receiver obtains its decryption key from a central authority, which needs to be trusted as it generates secret keys for every user.

IBE was proposed by Adi Shamir in 1984.he was however only able to give an instantiation of identity based signatures. Identity based encryption remained an open problem for many years .The pairing based Boneh-Franklin scheme and cock's encryption scheme based on quadratic residues both solved the IBE problem in 2001.

IBE system allows any party to generate a public key from a known identity value such as an ASCII string .A trusted third party, called the private key generator ,generates the corresponding private keys .To operate ,the PKG first publishes a mater public key and retains the corresponding master private key. Given the master public key ,any party can compute a public key corresponding to the identity ID by combining the master public key with the identity value .To obtain a corresponding private key the party authorized to use the identity ID contacts the PKG ,which uses the master private key to generate the private key identity ID.

As a result, parties may encrypt messages with no prior distribution of keys between individual participants. This is extremely useful in cases where Pre distribution of authenticated keys is inconvenient or infeasible due to technical restraints. However, to decrypt or sign messages, the authorized user must obtain the appropriate private key from the PKG.

2. LITERATURE REVIEW

Identity based encryption is an arousing curiosity modification for public key encryption, it is generated to compute the key management in a certificate based public key infrastructure using user information such as name email ID , address etc and it is considered as public key. So that we can tell that, public key and certificate is not necessary for the sender using IBE, But using a receiver identity we can directly encrypt the message.

With the corresponding identity, receiver obtaining the private key from the private key generator and PKG is used to decrypt the ciphertext.

An arbitrary string is used as public key for the IBE, is represented as an advantages of against PKI, which demands for the accurate revocation technique.

Server revokes the users from the system when the private keys of user get expired. Revocation is based on validity period for certificates

Boneh and Franklin are the first implementers, IBE was researched by using a cryptographic application.

Security was proven in random oracle for the first model. Under the selective ID and adaptive ID some subsequent systems are achieved provable secure, now a days for the IBE systems there have been number of lattice based constructions are presents. Boneh and Franlin's were given more suggestions but it is impractical.

Hanaoka et al. implements that, users should renew their own secrete keys periodically when without communicate with the PKG. in this work every user should possess a tamper resistant device. And here it has one more solution is that: mediator aided revocation means there is a trusted third party to supports users to decrypt the all ciphertext.

Lin et al. proposed A Space efficient revocable IBE techniques from ABE but it requires bilinear operations for a single decryption.

Libert and vergnaud improved the revocable IBE schemes and they focused on the enhancement of security.

3. PRELIMINARY

Cryptographic background:
Bilinear Map: e: $G_1 \times G_1 \rightarrow G_2$.

- G_1, G_2 are cyclic groups of same prime order p:
- G_1: Additively written. G_1=(p);
- G_2: Multiplicatively written.

Known examples: weil and tate pairings.
G_1: Subgroups of an elliptic curve group.
G_2: Subgroups of the multiplicative group of a finite field.
Bilinear Map: Properties
Bilinearity: e(aP,bP)=e$(p,p)^{ab}$.
Non –degeneracy: e(p,p)≠1.
DBDH: using of this advantages are
Let A be a probabilistic algorithm
Input: (P, P_1 ,P_2 ,P_3, Z) ∈ $G_1^4 \times G_2$;
Output: a bit b (denoted by =A→b).

4. IDENTITY BASED ENCRYPTION

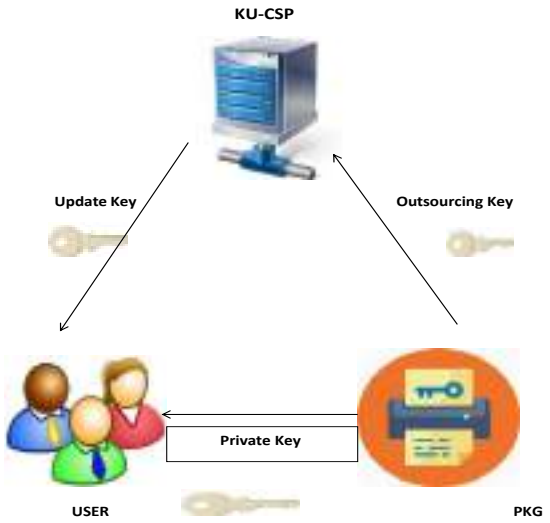


Fig.1. model of KU-CSP

Identity based encryption it consists of 4 algorithms that forms a complete IBE system:
Setup: this algorithm is run by the PKG one time for creating the whole IBE environment. The master key is kept secret and used to derive user's private keys, while the system parameters are made public .It accepts a security parameters k (i.e. binary length of key material) and outputs:
1.A set *P* of system parameters, including the message space and ciphertext space *M* and *C*.
2. A master key *Km*,
Extract: This algorithm is run by the PKG when a user requests his private key .Note that the verification of the authenticity of the requestor and the secure transport of *d* are problems with which IBE protocols do not try to deal .It takes as input *P,Km* and an identifier
ID ∈ $\{0,1\}^*$ and returns the private key *d* for user *ID*.
Encrypt: Takes *P*, a message $m \in M$ and $ID \in \{0,1\}^*$ and output the encryption $c \in C$.
Decrypt: Accepts *d,P* and $c \in C$ and returns $m \in M$.

5. PROBLEM STATEMENTS:

KU-CSP is used to realize the revocation for compromised users. KU-CSP can be used as a public cloud run by a third party to deliver basic computing capabilities to PKG as standardized services over the network. Revocation is triggered, instead of requesting private keys from PKG is unrevoked user.
How does Alice regain her privacy?
Basic idea: double encryption: combine a PKE and an IBE many subtleties to take care of.
The steps involved in the depicted in this diagram: Identity based encryption: Offline and Online.

Security definition:
Game between adversary and simulator.
Set-up: simulator

- Generates pp and master key.
- Provides the adversary with PP.
- Keeps master key secret.

Phase 1: adversarial queries.

- Key extraction oracle: ask for the key of my identity.
- Decryption oracle: ask for the decryption of any ciphertext on any identity.
- Restraction: cannot ask for decryption using ID, if a key ID has been asked earlier.

Challenge:

- Adversary outputs and two equal length message M_0 and M_1.
- Adversary should not have asked for the private key of ID^* .
- Simulator chooses a random bit b; encrypts M_b using ID^* to obtain C^*; Gives C^* to the adversary.

Phase 2: adversarial queries.

- Same as phase1.
- More restrications:
  Can't ask for the private key of ID^*;
  Can't ask for the decryption of C^* under ID^*.

Advantage: $\epsilon = 2 \times |P_r[b=b'] - 1/2|$.
(ϵ, t)-adversary: running time t: advantagesϵ.

- Storage definition:

Full model: adaptive –ID and CCA- secure.

- Weaker definitions:
- Adaptive –ID and CPA-secure.
  Adversary not provided with the decryption oracle.
- Selective –ID
  Adversary has to commit to the target identity even before the protocol is set up.
  CPA-secure.
  CCA-secure.

Revocable IBE and security: An identity based encryption with efficient revocation or simply revocable IBE scheme *RIBE=(S,SK,KU,DK,ϵ,D,R)* is defined by an algorithms and has associated message space M , identity space I and time space T. we assume that of T is polynomial in the security parameter. Each algorithm is run by either one of three types of parties-key authority, sender or receiver.

Key authority maintains a revocation list rl and state st. Revocation list rl can be a part of state st, but we keep it explicit for clarity.

- Stateful setup algorithm S(run by key authority) takes input parameter 1^k and number of users n, and outputs public parameter p^k,master key mk, revocation list rl(initially empty)and state st.
- The stateful private key generation algorithm SK(run by key authority)takes input public parameter pk, master key mk ,identity $\omega \in$ I and state st. and outputs private key sk_w and an update state st.
- The key update generation algorithm Ku(run by key authority) takes input public parameters pk ,master key mk, key update time t ∈ T, revocation list rl and state st, and outputs key update ku_t.

Security of revocable IBE

We define the selective revocable ID security for Revocable IBE scheme .our security model captures the standard notation of selective –ID security but it also takes into account possible revocations. Since we explicitly consider time period, in the beginning of the experiment in addition to the challenge identity the adversary also declares the challenge time. Just as in the standard selective –ID security definition the adversary can request to learn user's keys. In addition we let the adversary to revoke users of its choice at ant period of time and all key updates.

Unlike in the standard security model, we allow the adversary to learn the private key for the challenge identity, but only if it was revoked prior to or at the challenge time .the adversary given a ciphertext of one of the two messages of is choice encrypted for challenges identity and time. It has to guess which of the messages was encrypted.

First we define (selective) security against chosen –plaintext attack and then show how to extend the definition to chosen –ciphertext attack.

6. SYSTEM ARCHITECTURE

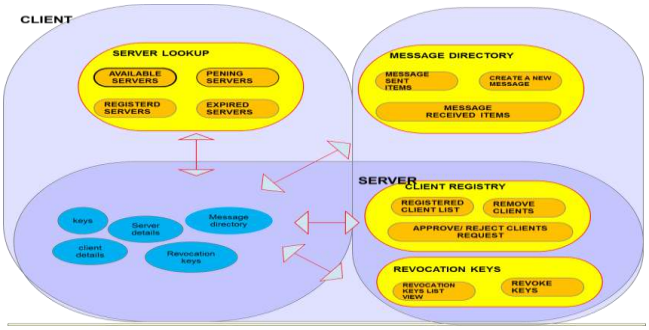


Fig.2. architecture diagram

Available servers: servers currently available for communication.

Register server: To preserve the connection information.

Expired server: whose register key has got expired.

Registered client list: list of clients who are registered with the server.

Remove client: remove the clients from the list whose key has got expired.

Approve/reject client request: approve or reject client request based on the request made by the client.

Client details: details of the server responses stored in the client.

Message sent items and message received items: number of and details of messages received and sent between client and server. details: details of the requests and messages sent by the client.

Revocation keys: Used to provide access privileges.

7. MAIN CONSTRACTION

In our IBE scheme message are encrypted for two "attributes": identity of the receiver and time period. The decryption key is also computed for attributes identity and time, on a first – degree polynomial, meaning both attributes of the decryption key must match with those of a ciphertext in order to decrypt.

We split the decryption key in two components corresponding to identity and time that we call private key and key update respectively.

The private key is issued to each user by the key authority just like regular private key in IBE. The key update is published by the authority and publicly available to all users.

8. PERFORMANCE ANALYSIS

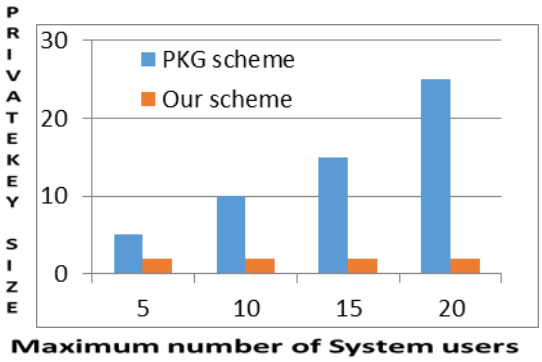


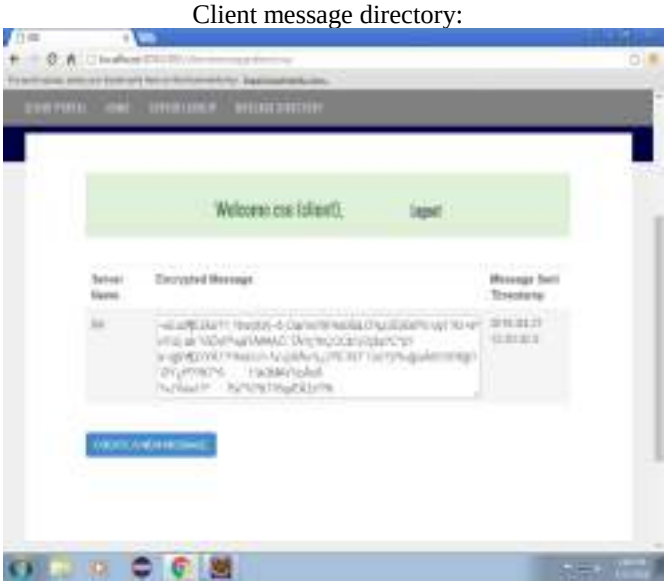
Fig.3. performance comparison

In this section, we will provide a thorough experimental evaluation of the construction proposed. In the above figure, we show the comparison on private key size. Besides the better performance in efficiency and private key size, another advantage of our scheme over the previous work is that it supports dynamic number of users. Specifically, the previous work requires to fix the maximum number of users in system initially to facilitate building the binary tree. Once the maximum number is fixed, it is difficult to add users exceeding this bound. Ours does not have such a drawback, and flexibly supports dynamic management of users.

9. SNAPSHOTS

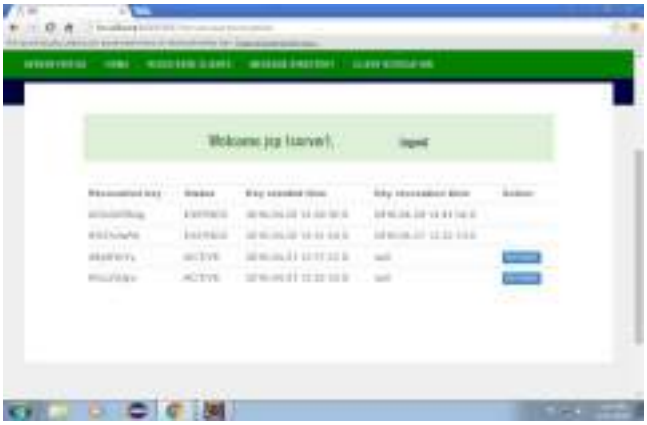
Client Registration:

Client message directory:



Server lookup:

Server user revocation:



10. CONCLUSION AND FUTUREWORK

Providin a critical issue of identity revocation, we introduce outsourcing computation into IBE and propose a revocable scheme in which the revocation operations are delegated to CSP. With the aid of KU-CSP, the proposed scheme is full-featured:

1) It achieves constant efficiency for both computation at PKG and private key size at user.

2) User needs not to contact with PKG during keyupdate, or, PKG is allowed to be offline after sending the revocation list to KU-CSP.

3)Nose cure channel or user authentication is required during key-update between user and KU-CSP. Finally,we provide extensive experimental results to demonstrate the efficiency of our proposed construction.

Furthermore, we consider realizing revocable IBE under a stronger adversary model. We present an advanced construction and show it is secure under RDoC model, in which at least one of the KU-CSPs is assumed to be honest. Therefore, even if a revoked user and either of the KU-CSPs collude, it is unable to help such user re-obtain his/her decryptability. Finally, we provide extensive experimental results to demonstrate the efficiency of our proposed construction.

ACKNOWLEDGMENT

We would like to thank Mangala.C.N, Associate Professor, Department of Computer Science and Engineering, East West institute of technology for her constant support and her guidance and also for providing us the opportunity to contribute on data access through users identity. At last we also thank all faculty members of Department of Computer Science and Engineering for helping us in this work.

REFERENCES

[1].M.Green,S.Hohenberger, and B.Waters, "*outsourcing the decryption of ABE ciphertexts*," in proc.20th USENIX Conf. security(SEC11),2011,PP.34-34

[2].Brent waters, "*ciphertext-policy attribute-based encryption* "An expressive,efficient,and provably secure realization. In PKC 53-70, 2011.

[3].Vipul Goyal, Omkat pandey,Amit Sahahi, and Abhishek jain."*Bounded ciphertext policy attribute based encryption*" in ICALP, 2008, PP 579-591
[4].John Bethencourt. "*Ciphertext-policy attribute based encryption library*". From http//acsc.cs.ut exas.edu ,2010.
[5].C. Delerablee, "*identity based broadcast encryption with constant size ciphertexts and private keys*" in cryptology-2007,ed springer,pp:200-215
[6].S.Honenberger and B.waters,"*Attribute based encryption with fast decryption of ABE ciphertext*", USENIX, 2011.
[7].S.Yu, C.Wang,K.Ren, W.Lou,"*Attribute based data sharing with attribute revocation*", security(ASIACCS'10),PP.261-270,2010

AUTHORS BIOGRAPHY

Mangala.C.N received the B.E degree in Computer science and Engineering from NCET, Bangalore, VTU in 2006 and got M.Tech degree in Computer Science from RVCE, Bangalore, India. She is currently working as Associate Professor in the faculty of CSE, EWIT- Bangalore, India. Her area of interest includes Image Processing, Data mining, Cloud computing and Big data.

Bhoomika M.U is pursuing her B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes cloud computing and Big Data, Data Mining.

Chandini M is pursuing her B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Cloud Computing, Big Data and Network Security.

Harshitha R S is pursuing her B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Cloud Computing, Big Data and Network Security.

Suraj S Gurav is pursuing his B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Cloud Computing, Big Data,Data Mining.

Be-Safe: Application to Limit Druken Drive Cases

Bharath B K, Abdul Rahoof K, Ashwin Suresh, K Rahul
Asst. Professor Department of CS&E,
Coorg Institute of Technology
rasrahoofr@gmail.com, ashwinsuresh2020@gmail.com,rahulkadavl@gmail.com

ABSTRACT-The amount of road accidents are increasing day by day, the main reason for that is drunk driving. Here wedeliver a solution for drunken driven accidents by implementing an application that will analyze and check whether the driver is fit to drive or not.

Keywords-Android, Arduino ,Blood Alcohol Content, Be-Safe and MQ3 Sensor.

I. INTRODUCTION

Smartphones are no longer a need but a style of living. They can be used for communication, social networking. Statistics showthat 51% of users in the world are using devices with Android OS. The application which we are implementing works in Android mobiles. According to recent surveys 1,3700 people were killed in road accidents that is more than number of people killed in all our wars put together[1].There is the one Demise every four minutes due to road accidents in India.

Drunken Driving is one of the leading casualties of road accidents. One serious road accidents in the country occur every minute and 16 die on Indian roads every hour [2].Around one-third of the drunken driving issues such as - arrests, crashes, deaths and injuries comes from repeat offenders. At any given points we potentially share the roads with two million people with 3 or more drunken driving offenses.

Alcohol can slow the pace of communication between neurotransmitter in the brain which carries messages between neurons. These chemicals can either intensify or minimize your body responses feelings and your mood [3].Alcohol shrinks and disturbs brain tissue which will lead to color-blindness and other effects.

In this application we provide you to check whether you're fit for driving after alcohol consumption. Various test like color identification, simple mathematical puzzles, horizontal movement and blood alcohol content sensing.

II. LITERATURE REVIEW

Android is a mobile operating system (OS) currently released by Google, built on the Linux kernel and intended primarily for touchscreen mobile devices such as smartphones and tablets. Android's user interface is mostlycenteredon straightoperation, using touch gestures that loosely correspond to real-world activities, such as swiping, touching and pinching, to manipulate on-screen objects, along with a virtual keyboard for text input.

Applications ("apps"), which extend the functionality of devices, are written using the Android software development kit (SDK) and, mostly the Java programming language which has complete access to the Android APIs. Java may be combined with C/C++, along with a choice of non-default runtimes that allow better C++ support, the Go programming language is also supported by laterversion from 1.4, which can also be used exclusively although with a restricted set of Android APIs. The SDK contains a complete set of development tools, including a debugger, software libraries, a handset emulator based on QEMU, documentation, sample code, and tutorials.

An Arduino board historically consists of an Atmel 8-, 16- or 32-bit AVR microcontroller with matchingconstituents that ease programming and incorporation into other circuits. An important feature of the Arduino is its normal connectors, which let users connect the CPU board to a variety of interchangeable add-on componentscalled shields. Certain shields interconnect with the Arduino board directly over various pins, but many shields are independently addressable via an I²C serial bus—so many shields can be stacked and used in parallel. Before 2015, AuthorizedArduinos used the Atmel megaAVR series of chips, specifically the ATmega8, ATmega168, ATmega328, ATmega1280, and ATmega2560. In 2015, components by other manufacturers were added. A handful of other processors have also been used by Arduinofriendly devices. Most boards include a 5 V linear regulator and a 16 MHz crystal oscillator ,whilecertainschemes such as the LilyPadworks at 8 MHz and dispense with the onboard voltage regulator due to specific form-factor limitations. An Arduino's microcontroller is also pre-programmed with a boot loader that simplifies uploading of codes to the on-chip flash memory, related with other devices that typically need an external programmer. This makes using an Arduino more direct by allowing the use of an ordinary computer as the programmer. Currently, opt bootloader is the default boot loaderset up on Arduino UNO.

Arduino programs may be written in any programming language with a compiler that yields binary machine code. Atmel provides a development environment for their microcontrollers. Arduino can be connected to android phone either using Bluetooth or OTG cable. Arduino is programmed to sense the value of alcohol content when a person blows on to the sensor. MQ3 sensor is used to sense the alcohol content in blood.

The MQ-3 is a heat-driven alcohol device that outputs an analog, which through the use of your Arduino code and calibration, can be interpreted for whatever use you need.MQ3 consists of 3 pins A,B and H. Both the A pins are electronically the identicalto the B pins. The midpoint pins on eithersides are the heater element pins. Since the circuit will be running on +5V DC it doesn't matter which way the device is soldered to the board.

The aim of this application is to check whether you're fit for driving after alcohol consumption. Various test like color identification, simple mathematical puzzles, horizontal movement and blood alcohol content sensing.

III. SCOPE

- Implementation of solution is going to be done on mobile device with Android OS.
- Application can be used to limit drunken driving accident cases.
- Basic tests are to be taken to check whether you're fit to drive or not.

IV. METHODOLOGY

This research aims to provide a solution to the drunken drive cases. The following steps were executed

The **First step** is the user registration, when app is installed the user has to register with their name, weight, gender and mobile number where the messages will be sent.

The **Second step**is done after alcohol consumption, the user has to identify and perform various test like simple mathematical problems, blood alcohol content calculation and horizontal movement.

The **Third step** is after taking all data from the test and sensor value calculated, application check whether a person is fit for driving or not and a particular message will be send to mobile number provided during registration along with the user location details such as latitude and longitude and Google Map link.

V. IMPLEMENTATION

The application developed has various steps, The beginning of the application user have to register by providing Name, Phone, Weight, Gender and Mobile Number where sms can be send. After the registration user has to identify the Red color from a series of color and click the button when text color is red. If he fails to identify red color then the person is not fit to drive. Similarly he has to select text with Green and Yellow Color.

The user location is identified by the application latitude and longitude will be calculated and stored once the test is done based on the results the SMS is send to the number provided along with user latitude and longitude and Google Map Link

Fig1Application Starting

The second step is alcohol sensing the android application can be connected to sensor based on Arduino and MQ3 sensors. The sensor is connected to Arduino and Arduino is programmed to read value of blood alcohol content when the user blows to sensor. The sensor is can be connected to Mobile phone either using Bluetooth or USB OTG cable.

Fig 2User Registration

Fig 4 Color Picking

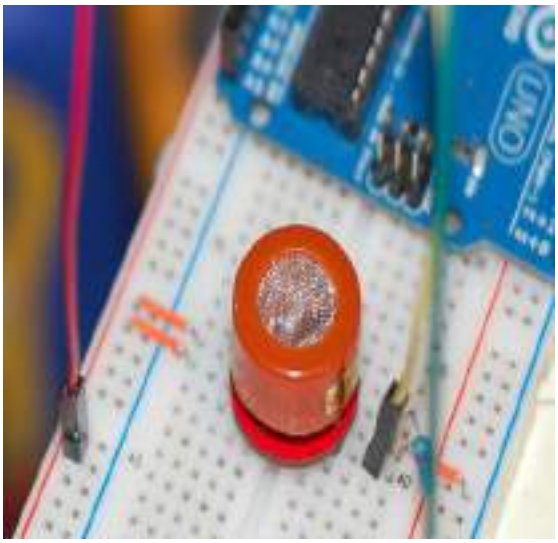


Fig 5Sensor

The third step is simple mathematical puzzles it's done to check the mental status whether they can deal with a particular situation when occurs. The user have to provide solution for simple mathematical question based on which the application determines whether you are fit to drive.

Fig 4 Puzzle

The fourth step is horizontal movement where user to have to move a certain distance in horizontal direction this test is used to check whether user mental stability is flawless.

The final step is based on the above four test application will calculate Blood alcohol content and with other statistics from the test application will predict whether you're fit to drive or not. If you're not fit a message will be send to the number provided during registration along with the user location latitude and longitude to the number provided.

VI. CONCLUSION

6.1 LIMITATION

- Is not for hefty drinkers.
- User may not provide proper data during registration and test.

6.2 FUTURE SCOPE

This is a prototype application and sensor. The application can be also made in such a way that it will lock the vehicle permanently if the user fail in test and it will be unlocked when he passes the test.

A SMS can be sent directlyto Police or other agencies about the location of the drunken driver so that they can take necessary actions.From this paper we like to present you an application which can be used to limit drunken drive accidents.

REFERENCES

[1] sites.ndtv.com/roadsafety/important-feature-to-you-in-your-car/

[2]National Crime records bureau, Ministry of road transport and highway, Law commission of India, Global status report on road protection.

[3]pubs.niaa.nih.gov/publications/Hangovers/beyondhangover.htm

On Survivability Testing and Computing the Node Connectivity of a Topological Structure of a Interconnection Network Graph

Nethravathi B[1], Kamalesh V N[2]
[1]*Department of ISE, JSSATE, Bangalore,* [2]*Department of Science and Engg., TJIT, Bangalore,*
[1]nethravathi.sai@gmail.com, [2]kamalesh.v.n@gmail.com

ABSTRACT- One of the major objectives in the design of topological structures of interconnection survivable computer communication networks is the construction of k-node connected networks. This is to achieve maximum survivability in the presence of link or node failure. Once a potential network structure is constructed by an existing algorithm or network is synthesized by merging other networks, and a claim is made that the generated network topology is k-node connected. There are few heuristics available in literature to verify the same. This research article reviews and analyzes the most popular methods and strategies for testing survivability of a topological structure of a interconnection network graph.

Keywords: physical network topology; topological design; fault tolerance; survivability; k-connected;

I INTRODUCTION

The topological design of an interconnection network structures specifies how the communication nodes of a network are interconnected by links. This is a significant stage in the design of survivable network structure, because the routing algorithm flow control, transmission delay etc., depends mainly on the designed network topological structures. [2][8]

The mathematical graph theory concepts and ideas are extensively used in computer science applications especially in the field of computer networks. [4]

The topological structure of network can be represented and mathematically modeled by using a simple graph where the vertices of the graph represent the processing components (nodes) of the network and the edge represents the communication links. On the other way, any simple graph represents a topological structure of some networks. Thus the mathematically graph models and the network topological structures are one and the same. [8][4]A linear graph is an order pair or sets G = (V,E), where set V = { v_1, v_2, …} whose elements are called vertices of the set E = {e_1,e_2 , … } whose elements are called the edges, such that each edge e_i is associated with a pair of vertices that is e_j = (v_i v_j).A simple graph is a graph with no parallel edges and loops. A graph with no loops is a multigraph.A path P in a graph G = (V,E) is an alternating sequence of distinct vertices and distinct edges starting and ending with vertices. A graph G = (V,E) is said to be connected if there exists at least one path between every pair of vertices otherwise it is said to be disconnected. A component of a graph G is a maximal connected sub graph. A disconnected graph has more than one component, whereas a connected graph has exactly one component.

Two vertices v_i,v_j in a graph G (V,E) are said to be adjacent if they are connected by an edge. Two edges are said to be adjacent, if they share a vertex. Degree of a vertex v in an undirected graph G = (V,E) is the number of edges incident on it. Degree D (G) of a graph G = (V,E) is the smallest degree of all the vertices in G.

D (G) = min{deg(v) | v in graph G }.

The vertex-connectivity (node connectivity) of a connected graph G is denoted by K (G). It is the minimum number of vertices whose removal can either disconnect G or reduce it to a 1-vertex graph.

The edge connectivity (link connectivity) of a connected graph G denoted by **λ**(G), is the minimum number of edges whose removal can disconnects G.

A vertex-cut in a graph G is a vertex set U such that G-U has more components than G. An edge-cut in a graph G is a set of edges D such that G-D has more components than G.

The relation between node connectivity, link connectivity and degree of a graph is given by Whitney's inequality and is as follows

$K(G) \leq \lambda(G) \leq D(G)$ **. . .**(1)

From (1) it is the node connectivity more crucial than the link connectivity because in equation (1) if K is more or higher, then definitely λ is always higher, that is, in particular, if survivable network can withstand the loss of p nodes, it can also definitely withstand the loss of p links.

If a graph is *k*-connected, there are k node disjoint paths between any two distinct nodes. The degree of survivability of a network improves ask is made larger and larger.

The survivability of a network is the fundamental issue in the design of interconnection network. The networks are survivable if their underlined network topological structure is fault tolerant. It is the ability of the network system to execute the specified algorithm correctly regardless of the hardware failure and the program errors. If the physical component of the network system fails then its function has to be performed by other components of the same network system.

It is essential to have proper yardstick or a mathematical abstraction to measure the effect of the fault tolerance. A deterministic graph theoretical based measure for fault tolerance is k- node connectivity of network.[1][2][3][8][18]

II EXPLORATION OF THE RESOURCE MATERIALS, EXISTING HEURISTICS AND METHODS

A network engineers and mathematicians have proposed few algorithms and heuristics for computation or testing of node connectivity of interconnection network graphs.

The earliest method in this field is due to kleitman [17]. This method is applicable for the large network graphs. It makes use of ${}^{r}c_2 + {}^{r}c_{(k-r)}$ verifications instead of ${}^{k}C_2$ verifications. The input for this method is the given network graph N with n nodes and the number k (node connectivity number). The algorithm gives the output whether the given network is k - connected or not. The method starts by choosing any node n_1 and checks whether there exist k node disjoint paths between n1 and every other nodes of the network graph N.

If there does not exist k-node disjoint path, then network is declared same that it is not k-connected else delete the node n_1 and all the links/branches incident to the node n1 and obtained the network graph $(N - \{ n_1\})$. Now choose any node n_2 for the network graph $(N - \{ n_1\})$ and check for the (k-1) node disjoint path between n_2 and every other nodes of $(N - \{ n_1\})$. If we get the required number of node disjoint path, continue the process, if not stop and declare the same, that the network is not k –connected. By repeating the above process k – times, we end up a network graph $(N - \{ n_1,n_2 \ldots n_k\})$. This graph should be 1 - connected graph if not, graph is not k-connected.Tarjan [16] in his article has presented a method for checking 2-connected networks using the technique of Depth First Search. It is a linear time algorithm and he has illustrated the same.Hopcraft and Tarjan [15] have presented a technique for checking 3-node connectedness of a given network graph by dividing the network graphs in to triconnected components. It is a linear time algorithm. He proves that the method is theoretically optimal to within a constant factor and efficient in practice.S. Even and Tarjan [14] has described a model for the maximal stationary flow using the algorithm of Dinic by assuming that the capacities of node and links as one. Further as an application of the above result, they calculate the node connectivity of the network graph.S. Even [13] has presented an algorithm for testing whether the node connectivity for large network graph N with n number of nodes is at least k. This method works for undirected graph. He has also shown the above method also works for the directed graph with variance in algorithm and it starts by identifying the nodes of the network graph by using the counting natural numbers starting from

This algorithm has two phases. In the first phase; the first k nodes are checked for k disjoint paths and in the second phase, the new node X is augmented to original network by connecting X to every node of specially considered set L and check for k disjoint path between X and j where j Є L. For any j if there are less than k disjoint path, the network is not k connected, else k connected. This is repeated for all j belongs to L.

Becker[12] adds a probabilistic variant to the Even & Tarjan method with the minimum error probability in expected time for sparse group with the condition

$$P_r (\mu > k) \leq \varepsilon,$$

Where μ = min N (a,b) and N (a,b) = min ($|S| \leq V-\{a,b\}$ is an (a,b) vertex separator).

Esfahanian and Hakimi [11] presented a technique for computing k-edge connectivity of a network graph or a digraph with n/2 calls to the maxflow algorithm. Using this they have described a method to determine the given network is at least k-node connected by using n-k+1/2(k-1)(k-2) calls.

Kanevsky and Ramachandran [10] have presented a sequential algorithm to check the network graph for 4-node connectivity based open ear decomposition. Each vertex of the network graph is deleted and the resultant network in turn tested for 3-connectivity.

Joseph cherrian *et.al* [7] has presented randomized approach based deterministic approach testing methods for K vertex connectivity for a directed graph using Monte Carlo algorithm.

M. R. Henzinger and Rao[9] have presented deterministic algorithm to compute the node connectivity and finding the corresponding separator of the network graph. The main algorithm is the generalization of the previous preflow-push algorithm for network flow.

S.K Srivatsa et al [6] have presented a method to compute the node connectivity of a network graph by using the adjacency matrix, in their study of survivability of high speed topological network structures. It is a non iterative algorithm. The algorithm begins by numbering the node by appropriate numbering technique. Subsequently node disjoint paths between given pair of nodes is calculated by using the adjacency relationship between the pair of nodes by using the matrix data structures. In their methods the nodes of graph are partitioned in to 2 sets L_1 and L_2, and the corresponding counters are created. Using the adjacency relationship counters are updated. Once all the nodes corresponding counter values are checked for adjacency then the values of the counters are sorted. The lowest value of the sorted counter list gives the node connectivity number of the network graph.

III. ALGORITHMIC PROFILE ANALYSIS, RESULTS AND DISCUSSIONS

The heuristics, algorithms and methods presented by various authors on Survivability testing and computing the node connectivity of a topological structure of an interconnection network graph are studied in detail. A theoretical profiling of the well known methods is assessed by detailed examination.

Based on the results, in this section we portray the pros and cons of the above proposed heuristics and algorithms are discussed below in brief.

The Kleitman [17] method is the oldest method and is widely used by mathematicians and early network engineers. This method just checks whether the given network is k-connected or not.

The methods due to Tarjan [16], Hopcroft & Tarjan [15], Kanevsky & Ramachandran [10] are very specific methods

applicable only to check for 2-node connectivity, 3-node connectivity and 4-node connectivity respectively.

The methods due to Kleitman [17] and S.Even [13] are the methods which just tests or checks for the given k-connectivity, *i.e*. it takes the network graph N (V,E) the number of nodes n and the connectivity number k and the expected output is given network is k-connected or not. To calculate or compute connectivity of the network graph one has to execute the algorithm recursively for k=1, 2, 3…

However the method due to Henzinger & Rao [9], Esfahanian & Hakimi [11] and S K Srivatsa [6] are generalized methods for computing the node connectivity of a given network that is these algorithm takes the network graph N (V,E) as an input and output the corresponding node connectivity k. Esfahanian & Hakimi [11] is based on max flow algorithm and applicable for both undirected and directed network graph. The method due to Henzinger & Rao [9] is a randomized and deterministic algorithm.

The method due to joseph cherrian *et al* [7] is applicable for digraphs. In reality, these methods applicable only for simplex mode communication.

S K Srivatsa [6] method is a direct and generalized method which makes uses of one dimensional and two dimensional data structure for programming. The method takes the entire network details as input to the algorithm and output the connectivity number k. This method is applicable and suitable for the design of small and medium network with optimal number of closed circuits passing through a single node.

The detailed analysis of the above work is tabulated in table1.

IV. CONCLUSION

This research article has presented the explorative and comprehensive review of the existing methods and recent developments on survivability testing and computing the node connectivity of a topological structure of a interconnection network graph. In the analysis the methods are classified as iterative and non iterative approaches applicable for either testing or computing. It is observed from our analysis that the previous methods are restricted only to either checking the node connectivity or testing. However there is no single method which attempts to address both completely, hence there is a scope for further research.

REFERENCES

[1] Santanu Saha Ray, 2013, *Graph Theory with algorithms and its applications*, First Edition, Springer Publishers.

[2] Andrew S. Tanenbaum, 2011, *Computer networks*, 5th Edition, pearson education.

[3] Maarten van Steen, *An Introduction to Graph Theory and Complex Networks*, Copyrighted material - January 2010.

[4] S.G.Shirinivas, S.Vetrivel, Dr. N.M.Elango, *Applications of graph theory in computer science an overview*, International Journal of Engineering Science and Technology Vol. 2(9), 2010.

[5] Abdol-Hossein Esfahanian *On the Evolution of Graph Connectivity Algorithms*, Computer Science and Engineering Department, Michigan State University. East Lansing, Michigan. http://www.cse.msu.edu/eesfahani/book chapter/ Graph connectivity chapter.pdf

[6] Kamalesh V. N. and Srivatsa S. K. **(**2010**),** *Adjacency matrix based method to compute the node connectivity of a computer communication network*, Int. Journal of Computer Science Issues, vol. 7, pp. 62-67.

[7] Joseph Cheriyan and John H. Reift, 2005, Vol. 14, No. 4, pp. 435-451, *Directed s-t numberings, rubber bands, and testing digraph k-vertex connectivity*, Springer Berlin / Heidelberg.

[8] Junming Xu, 2004, *Topological structure and analysis of interconnection networks*, Second Edition, Kluwer Academic Publishers.

[9] M. R. Henzinger, S. Rao and H. N. Gabow, *Computing vertex connectivity: new bounds from old techniques*, Proc. 37th IEEE F. O. C. S. (1996), 462— 471.

[10] Kanevsky and V. Ramachandran, *Improved algorithms for graph fourconnectivity*, J. Comp. System Sci 42 (1991), 288—306.

[11] H. Esfahanian and S. L. Hakimi, *On computing the connectivities of graphs and digraphs*. Networks (1984), 355-366.

[12] M. Becker, W. Degenhardt, J. Doenhardt, S. Hertel, G. Kaninke, and W. Keber, *A probabilistic algorithm for vertex connectivity of graphs*. Inf. Proc. Letters 15 (1982), 135-136.

[13] S. Even, *An algorithm for determining whether the connectivity of a graph is at least k*, SIAM J. Computing 4 (1975), 393-396.

[14] S. Even and R. E. Tarjan, *Network flow and testing graph connectivity*, SIAM J. Computing 4 (1975), 507-518.

[15] J. Hopcroft, R.E. Tarjan, *Dividing a graph into triconnected components*. SIAM J. Comput 2 (1973), 135-158.

[16] R. E. Tarjan, *Depth first search and linear graph algorithms*. SIAM J. Comput 1 (1972), 146-160.

[17] D. J. Kleitman, *Methods for investigating connectivity of large graphs*, IEEE Trans. Circuit Theory CT16 (1969), 232-233.

[18] Kamalesh V.N, PhD thesis , *Topological Design of Networks using Graph Theoretical Algorithms*, 2010.

[19] Nethravathi B, Kamalesh V N, Manjunath Guru V.G, *Topological Design of Computer Communication Network Structures : A Comprehensive Review*, accepted for publication at Indian Journal of Science and Technology.

SECURE SYSTEM WITH HIGHER RELIABILITY AND CONFIDENTIALITY USING DISTRIBUTED DEDUPLICATION OF MULTIMEDIA

[1]Aishwarya R,Anusha G Bedekar, Chandana R, [2]Dr. Malathy M
[1]Department of CSE,RRCE,Bangalore-74.
[2]Professor, Department of CSE,RRCE,Bangalore-74.
[1]ash.ramchandra@gmail.com,anushabedekar94@gmail.com,chandana.r13@gmail.com, [2]anandnmalathy@gmail.com

ABSTRACT-Data deduplication is for removing duplicate copies of data, and is used in cloud storage to reduce space and upload files which reduces bandwidth. There is only 1 copy for every file, image file, audio file or a video file kept in cloud although such a file is owned by an enormous range of users. Deduplication system improves storage utilization while reducing reliability. When data are outsourced by users to cloud, the challenge of privacy for data arises. To overcome the protection challenges, this paper commit to formalize the notion of distributed reliable deduplication system. we tend to propose new distributed deduplication systems with higher dependability within which the data parts are distributed across many different cloud servers. The data confidentiality and tag consistency are achieved by employing a deterministic secret sharing scheme in distributed storage systems, rather than victimization convergent encryption. Security checks verifies that our deduplication systems are secure in terms of the definitions.

Keywords—Deduplication, distributed storage system, reliability, secret sharing

.I. INTRODUCTION

Deduplication methods are mainly applied to standby data and reduce network and loading overhead by detecting and eliminating redundancy among data. Deduplication rejects redundant data by custody only one physical copy and revealing to any other copy , and by not maintaining multiple data copies with the same content. A number of deduplication systems have various deduplication such as client-side or server-side deduplications, file-level or block-level deduplications. With the wide-ranging usage of cloud storage, data deduplication techniques become more good-looking and critical for the management of very large amount of records in cloud storage services which motivates enterprises and administrations to farm out data storage as analyzed in many real-life case studies [1] According to the analysis report of IDC, the quantity of data within the world is predicted to succeed in forty trillion gigabytes in 2020. [2] Drop box, Google Drive and Mozy, are applying deduplication to avoid wasting the network bandwidth and therefore the storage price with client-side deduplication. Two kinds of deduplication are there in terms of the size: (i) file-level deduplication, that discovers redundancies between completely different files and removes these redundancies to scale back capacity demands, and (ii) block-level deduplication, that discovers and removes redundancies between information blocks. The file will be divided into smaller fixed or variable-size blocks. Fixed size blocks simplifies the computations of block boundaries, whereas using variable-size blocks (e.g., supported Rabin fingerprinting) [3] Provides better deduplication efficiency. It also reduces the reliability of the system. A very crucial issue in a deduplication storage system is data reliability because there's only 1 copy for every file hold on within the server shared by all the owners. If a shared File/chunk was lost, quite a large amount of data becomes inaccessible. this is often because of the inaccessibility of all the files that share this file/chunk. A essential downside is to ensure high data reliability in deduplication system. several of the previous deduplication systems have solely been considering only one server. Users and applications intend for higher reliability, particularly in archival storage systems wherever data are critical and will be preserved over lasting periods. Thus, the deduplication storage systems should offer reliability comparable to alternative systems.

I. II. RELATED WORKS

Gantz J & Raiusel D have projected, the "digital universe" — a measure of all the digital knowledge created, replicated, and consumed throughout one year. it's also a projection of the size of that universe to the tip of the last decade. The digital universe is made of pictures and videos on mobile phones uploaded to YouTube, digital cinemas immigrating the pixels of our high-definition televisions, banking knowledge swiped in an ATM, security footage at airports and major events just like the Olympic Games, subatomic collisions recorded the massive hadron collider at CERN, transponders recording road tolls, voice calls zipping through digital phone lines, and texting as a widespread implies that of communications. However, duplication among the file since cloud is used [2].

M. O. Rabin have projected, an information dispersal algorithmic program (IDA) is developed that breaks a file F of length L = (F (into n pieces F, 1 5 i 5 n, every of length (F, 1 = L/m, so each m pieces fulfil for reconstructing F. dispersal and reconstruction are computationally efficient. The total of the lengths (F, 1 is (n/m) . L. Since n/m are typically chosen to be close to I, the IDA is area efficient. IDA has varied applications to secure and reliable storage of data in computer networks and even on particular disks, to fault-tolerant and well-organized broadcast of data in webs, and for parallel processors to communicate with one

another. For the latter downside provably time-efficient and intensely fault-tolerant routing is achieved, using merely constant size buffers [9].

M. Bellare, S. Keelveedhi, and T. Ristenpart have projected, Cloud storage service providers like Dropbox, Mozy, et al. perform deduplication to avoid wasting space by solely storing one copy of every file uploaded ought to clients conventionally encrypt their files, however, savings are lost. Message-locked encryption (the most outstanding manifestation of that is convergent encryption) resolves this tension. However, it's inherently subject to brute-force attacks which will recover files falling into a notable set. we tend to propose an architecture that has secure deduplicated storage resisting brute-force attacks, and understand it in a system known as DupLESS. In DupLESS, clients encrypt beneath message-based keys via an oblivious PRF protocol. It allows clients to store encrypted data with an existing service and make them perform deduplication on their behalf, and nevertheless achieves robust confidentiality guarantees. We tend to show that encryption for deduplicated storage can achieve performance and space savings near that of using the storage service with plaintext data. it's an occasional performance thus doesn't match acceptable requirements fully [5].

M. Bellare has proposed new cryptographic encryption is formalized referred to as Message bolted encryption (MLE), where the key below which the encoding and cryptography is performed is itself derived from the message. MLE provides a replacement way to accomplish secure deduplication, a goal targeted by varied cloud storage providers. It provides definitions each for privacy and for a sort of integrity that is referred to as tag consistency. Supported this foundation each practical and theoretical contributions are created. Users would possibly wish their files to be encrypted and conventional encryption makes deduplication unfeasible [6].

J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou have proposed the basic idea of eliminating duplicate copies of storage data and limit the injury of purloined data if we've a tendency to decrease the value of that purloined data to the attacker. This paper attempts to address the matter of achieving efficient and reliable key management in secure deduplication. We have a tendency to initially introduce a baseline approach among which each {and every} user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates a large vary of keys with the increasing vary of users and wishes users to dedicatedly defend the master keys. To this end, we tend to propose Dekey, User Behavior profiling and Decoys technology. Dekey new constructions among those users haven't got to be compelled to manage any keys on their own but instead firmly distribute the convergent key shares across multiple servers for insider offender. As a symbol of idea, we tend to implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs restricted overhead in realistic environments. User profiling and decoys, then, serve a pair of functions. Initial one is validating whether or not data access is permitted once abnormal data access is detected, and other is that confusing the attacker with phony data. We have a tendency to posit that the mixture of those security measures can offer unprecedented levels of security for the deduplication in insider and outsider offender. The varied styles of data for each user hold on among the cloud and so the demand of long-term continuous assurance of their data safety, the matter of supportive correctness of data storage among the cloud becomes even more durable. Cloud Computing is not merely a third party data warehouse. The data hold on among the cloud may even be frequently updated by the users, still as insertion, deletion, modification, appending, reordering, etc. One crucial challenge of today's cloud storage services is that the management of the ever-increasing volume of data. Every user ought to associate an encrypted convergent key with each block of its outsourced encrypted data copies, thus on later restore the data copies. Although completely different users may share identical data copies, they have to possess their own set of convergent keys thus no various users will access their files. Second, the baseline approach is unreliable, as a result of it wants each user to dedicatedly defend his own master key. If the master key is accidentally lost, then the user data can't be recovered; if it's compromised by attackers, then the user data are leaked [11].

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller have proposed - As the world moves to digital storage for depository functions, there is an increasing demand for systems which can provide secure data storage in an exceedingly cost-efficient manner. By distinctive common chunks of data every among and between files and storing them one time, deduplication can yield cost savings by increasing the utility of a given amount of storage. Sadly, deduplication exploits identical content, whereas encryption makes a trial to make all content appear random; identical content encrypted with a pair of completely different keys results in very different cipher text. Thus, combining the space efficiency of deduplication with the secrecy aspects of encryption is problematic. We've developed a solution that has each data security and space efficiency in single-server storage and distributed storage systems. Encryption keys are generated throughout a uniform manner from the chunk data; so, identical chunks will perpetually encrypt to an identical cipher text. Furthermore, the keys can't be deduced from the encrypted chunk data. Since the data each user should access and decrypt the chunks that compose a file is encrypted using a key known exclusively to the user, even a full compromise of the system cannot reveal that chunks are used by that users. It'd have security issues [20].

M. Li, C. Qin, P. P. C. Lee, and J. Li have proposed cloud-of-clouds storage exploits diversity of cloud storage vendors to provide fault tolerance and avoid vendor lock-ins. Its inherent diversity property in addition permits us to produce keyless data security via dispersal algorithms. However, the keyless security of existing dispersal algorithms depends on the embedded random data that breaks data deduplication of the distributed data. To simultaneously enable keyless security and deduplication, we've an inclination to propose a novel dispersal approach referred to as convergent dispersal that replaces original random data with deterministic cryptographic hash data that is derived from the initial data

but can't be inferred by attackers whereas not knowing the complete data. We've a tendency to develop a pair of convergent dispersal algorithms, specifically CRSSS and CAONT-RS. Our analysis shows that CRSSS and CAONT-RS provide complementary performance benefits for varied parameter settings [16].

II. III. PROPOSED SYSTEM

In this paper, we analyze how to develop a secure deduplication systems with higher reliability in cloud computing. We introduce the concept of distributed cloud storage servers into deduplication systems to provide well error acceptance. In addition to shield data secrecy, the concept of secret sharing method is utilized, which is also well-matched with the distributed storage systems. A file is first divided and encoded into fragments by using the technique of secret sharing. These parts will be distributed across multiple independent storage servers. Furthermore, to support deduplication, short cryptographic hashtag of the content will also be computed and sent to each storage server. Only the data owner who leading to uploads the data is essential to use and issue the fragments, while all succeeding users who own the similar file copy do not want to do the similar task. To recover data copies, users need access a least number of storage servers through certain safety checks similar to authentication and obtain the secret shares. In other words, the secret shares of data can be accessible only by the users who own the corresponding data copy after being authorized. Four new safe deduplication structures are used to care deduplication with high consistency for file-level and block-level deduplication, individually. [1]The secret splitting technique, instead of old-style encryption approaches, is exploited to keep data privacy. Data are split into parts by using secure secret sharing schemes and stored at different servers.

III. IV. ARCHITECTURE

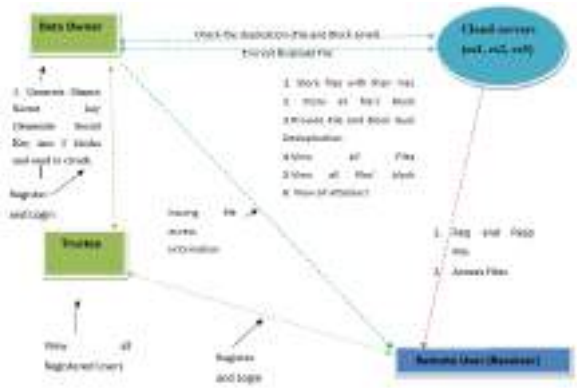


Fig 1:System Architecture

Data Owner

Here, the owner of the data uploads their data in the cloud server. [2]For the security purpose the data owner encrypts the file and then store in the cloud. The data owner can check the duplication of the file and multimedia file over Corresponding cloud server. The Data owner can have capable of manipulating the encrypted data file and the data owner can check the multiple cloud data as well as the duplication of the specific file. And also he can create remote user with respect to registered cloud servers. And also data owner has migrate to another cloud option, by this he can migrate files from one cloud server to another cloud server.

Trustee

In this module, the connector helps to check duplication of file existed or not in cloud server and you can check in multi cloud servers also. If it is existed, then also owner trying to upload the same file in same cloud server then connector automatically blocks his access permission. If it is not existed, then data owner can upload file in multi cloud servers at a time.

Cloud Server

[3]Data owners encrypt their data files and store them in the cloud for sharing with Remote User. [4]To access the shared data files, end users download encrypted data files of their interest from the cloud and then decrypt them.

Remote User

In this module, remote user logs in by entering his user name with the password. He will then request the cloud server for a secret key of the particular file. After getting secrete key he is trying to download file by entering file name and secrete key from cloud server.

Attacker Module

In remote user module, while downloading time if remote user entered any wrong file name or secrete key then cloud servers treats him as attacker and moves his access permission to block/attacker list.

V. IMPLEMENTATION

Fig 2: Owner login

Fig 3: Trustee

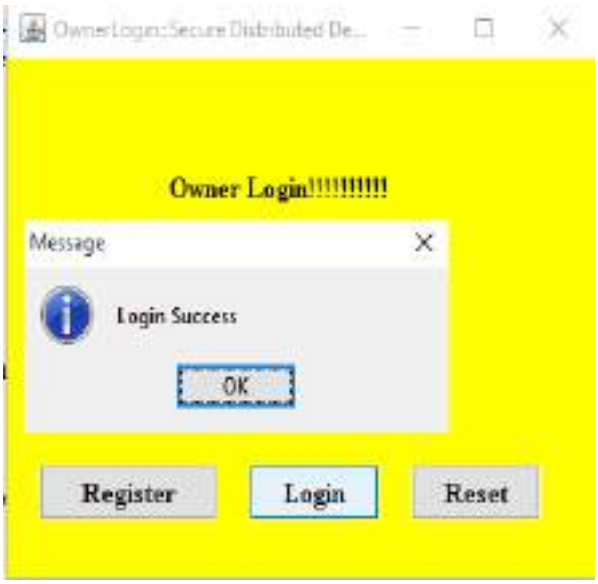


Fig 4: login successful

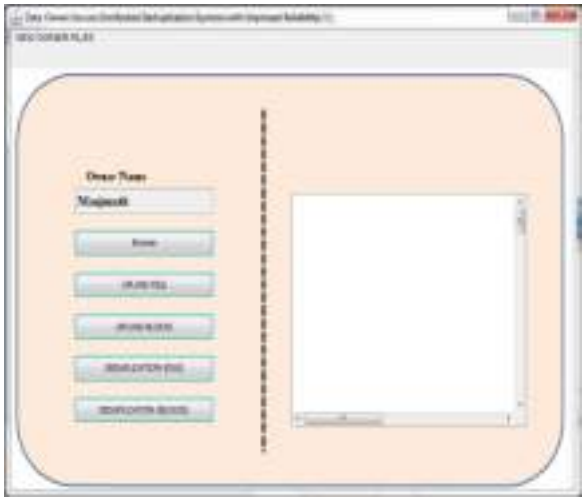


Fig 5:Data owner

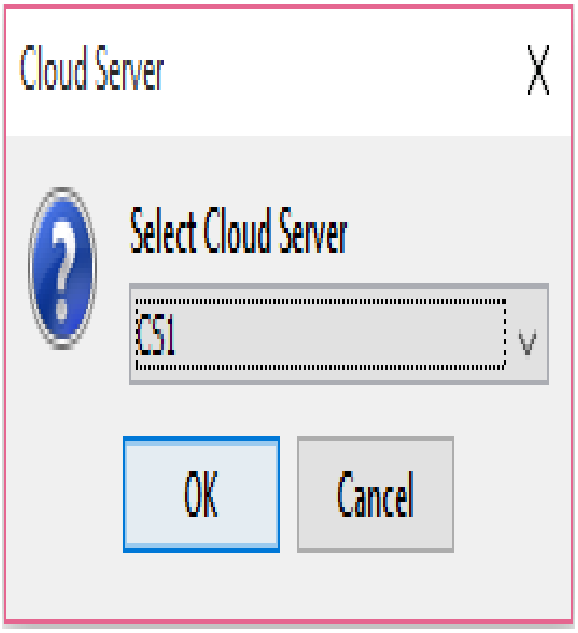


Fig 6: Select cloud server

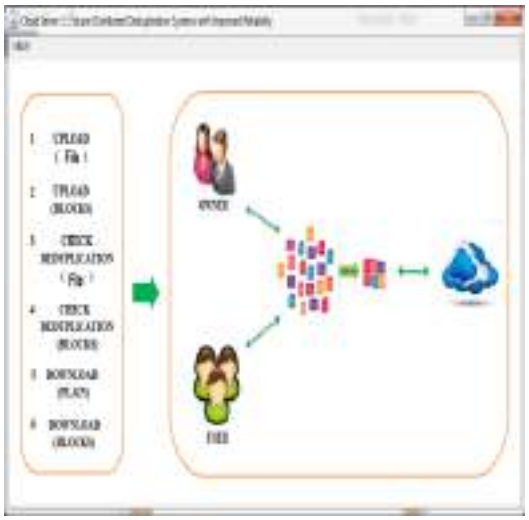


Fig 7:Cloud Server

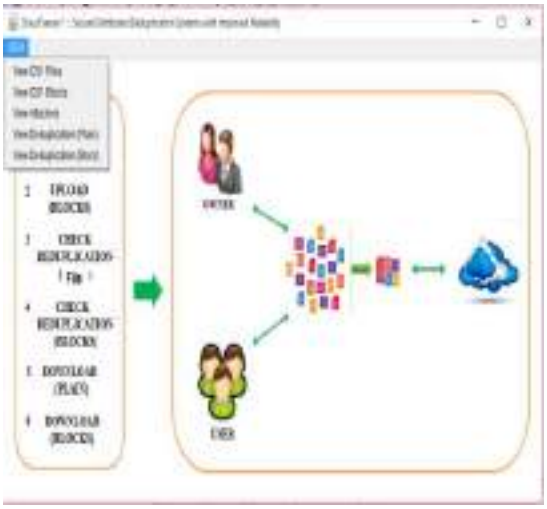


Fig 8: view files

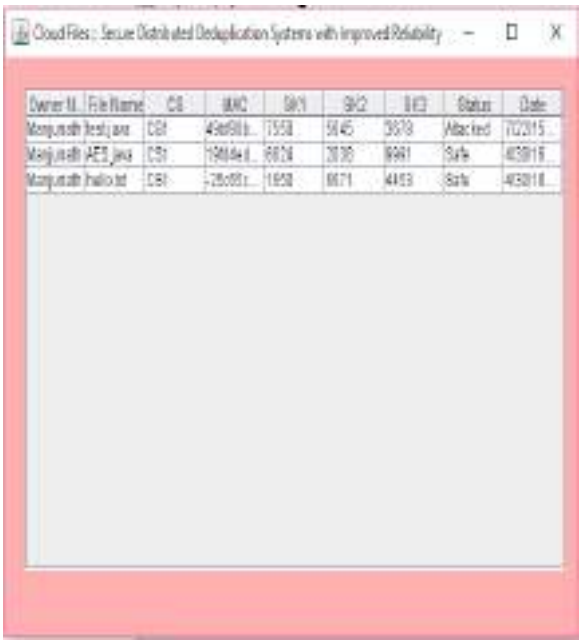


Fig 9: Files

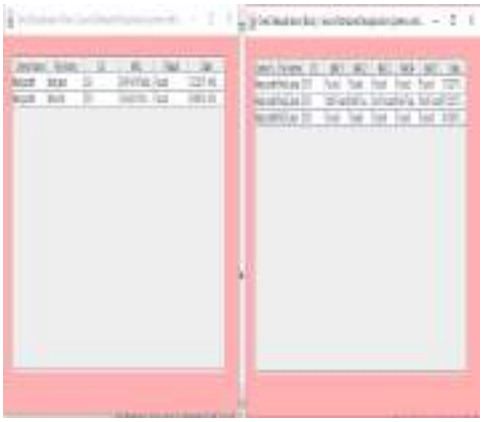


Fig 10: Check duplication

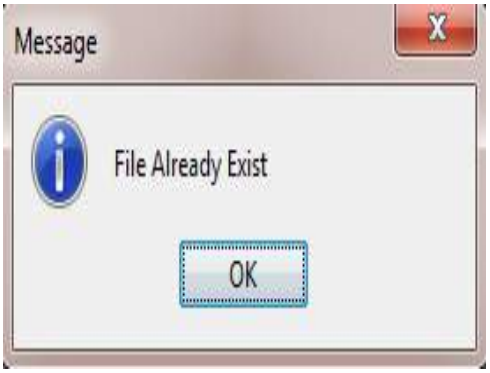


Fig 11: Pop up window

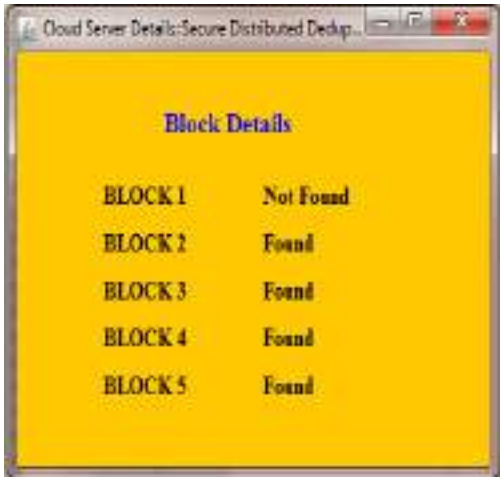


Fig 12: Block details

Fig 13: User login

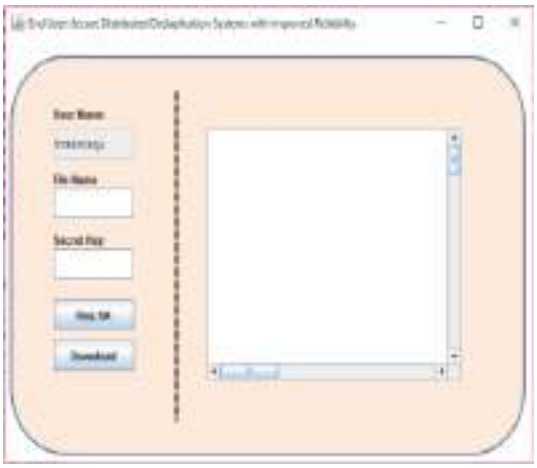


Fig 14: End user

Fig 15: secret key for download

VI. CONCLUSION

We proposed the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the users outsourced data without an encryption mechanism. Four constructions were proposed to support file-level and fine-grained block-level data deduplication. The security of tag consistency and integrity were achieved. We implemented our deduplication systems using the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.

REFERENCES

[1] Amazon,"CaseStudies," backup.
[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," http://www.emc.com/collateral/analyst-reports/idcthe-Digital-universe-in-2020.pdf, Dec 2012.
[3] M. O. Rabin, "Fingerprinting by random polynomials," Center for Research in Computing Technology, Harvard University, Tech. Rep. Tech. Report TR-CSE-03-01, 1981.
[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a

serverless distributed file system." in *ICDCS*, 2002, pp. 617–624.
[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in *USENIX Security Symposium*, 2013.
[6] M. Bellare,"Message-locked encryption and secure deduplication," in EUROCRYPT, 2013, pp. 296–312.
[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," in *Advances in Cryptology: Proceedings of CRYPTO '84*, ser. Lecture Notes in Computer Science, G. R. Blakley and D. Chaum, Eds. Springer-Verlag Berlin/Heidelberg, 1985, vol. 196, pp. 242–268.
[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1720–1728, Jul. 1999.
[9] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," *Journal of the ACM*, vol. 36, no. 2, pp. 335–348, Apr. 1989.
[10] A. Shamir, "How to share a secret," *Common. ACM*, vol. 22, no. 11, pp. 612–613, 1979.
[11] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in *IEEE Transactions on Parallel and Distributed Systems*, 2014, pp. vol. 25(6), pp. 1615–1625.
[12] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems." in *ACM Conference on Computer and Communications Security*, Y. Chen, G. Danezis, and V. Shmatikov, Eds. ACM, 2011, pp. 491–500.
[13]J. S. Plank, S. Simmerman, and C. D. Schuman, "Jerasure: A library in C/C++ facilitating erasure coding for storage applications - Version 1.2," University of Tennessee, Tech. Rep. CS-08-627, August 2008.
[14]J. S. Plank and L. Xu, "Optimizing Cauchy Reed-solomon Codes for fault-tolerant network storage applications," in *NCA-06: 5th IEEE International Symposium on Network Computing Applications*, Cambridge, MA, July 2006.
[15] C. Liu, Y. Gu, L. Sun, B. Yan, and D. Wang, "R-admad: High reliability provision for large-scale de-duplication archival storage systems," in *Proceedings of the 23rd international conference on Supercomputing*, pp. 370–379.
[16] M. Li, C. Qin, P. P. C. Lee, and J. Li, "Convergent dispersal: Toward storage-efficient security in a cloud-of-clouds," in *The 6th USENIX Workshop on Hot Topics in Storage and File Systems*, 2014.
[17] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in *Proc. of USENIX LISA*, 2010.
[18] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: the least-authority filesystem," in *Proc. of ACM StorageSS*, 2008.
[19] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, "A secure cloud backup system with assured deletion and version control," in *3rd International Workshop on Security in Cloud Computing*, 2011.
[20] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in *Proc. of Storages*, 2008.

Efficient Schemes for Resource Allocation In the Cloud for Media Streaming Applications

Shreya Kulkarni, PragnaPrakash S, Shaila S, Mrs.Janaki.K
Department of Computer Science, RRCE,Bangalore-74
kulkarnis642@gmail.com, karur.janaki@gmail.com

ABSTRACT-A large number of users in the Internet have recently attracted by the media streaming applications. Due to advent of this bandwidth-intensive applications, with guaranteed QoS relying only on central resources at a media content provider it is economically inefficient to provide streaming distribution. media content providers(e.g., Video on Demand (VoD) providers) can use to obtain streaming resources that match the demand by the elastic infrastructure offered by the cloud computing. The amount of resources allocated (reserved) in the cloud are charged by the media content providers. Based on non-linear time-discount tariffs (e.g., Amazon Cloud Front and Amazon EC2) the existing cloud providers employ a pricing model for the reserved resources. This type of pricing scheme offers discount rates depending non-linearly on the period of time during which the resources are reserved in the cloud. In this case, the right amount of resources reserved in the cloud, and their reservation time are used to decided by the open problem such that the financial cost on the media content provider is minimized. A simple - easy to implement- algorithm are proposed for resource reservation that maximally exploits discounted rates offered in the tariffs, while ensuring that as much as resources are reserved in the cloud. Our algorithm is carefully designed based on the prediction of demand for streaming capacity to reduce the risk of making wrong resource allocation decisions. The results show that the proposed algorithm significantly reduces the monetary cost of resource allocations in the cloud as compared to other conventional schemes according to our numerical evaluation and simulation.

Index Terms—Media streaming, Cloud Computing, Non-linear pricing models, Network economics.

1 INTRODUCTION

Huge number of users in the Internet is attracted by the media streaming application. As compared to the past few years as the number of video streams served increased 38.8% to 24.92 billion. Due to this huge demand it creates a burden on centralized datacenters at media content providers such as Video on-Demand (VoD) providers to sustain the required QoS guarantees [2]. The problem becomes critical with the increasing demand for higher bit rates required for the growing number of higher-definition video quality desired by consumers. In this paper, by using cloud computing we explore new approaches that mitigate the cost of streaming distribution on media content provider. A non-linear pricing scheme offers discount rates depending non linearly on the period of time during which the resources are reserved in the cloud. In this case, an open problem is to decide on both the right amount of resources reserved in the cloud, and their reservation time such that the financial cost of the media content provider is minimized. Based on the prediction of this (PBRA), of demand for streaming capacity, our proposed algorithm is carefully designed to reduce the risk of making resource allocation decisions. Cloud computing (IT infrastructure provided over the internet and priced over the usage).cloud computing here is used to avoid the unnecessary capital investment by the media content for unused data. Allocate memory on demand and optimize the resources. To target the expected usage peak, media content provider can make long term investment in infrastructure. This causes problem in view of flash crowd events. Expenses charged by cloud providers convert the upfront infrastructure investment for media content providers. Cloud service providers provide services over the internet and it is priced as it is used.Paas, Iaas, Saas are the services provided by cloud service providers. Where Iaas includes processor and can be scaled up as demand increases and Paas provides operation services from available system of operating system. Saas is used to select required software of your choice like Amazon, Microsoft and Google. These provide cloud services over the internet. Cloud enhances the auto scaling ability to reserve amount of resources to match the demand. Media content provider provides on demand plan and avoids unnecessary amount to be paid and also eliminate the waste by optimizing the space allocated in the memory. The on demand plan is the pay-per-use plan.Our main aim is to do in this paper is to reduce the monetary cost to reserved amount of resources in the cloud by implementing prediction based resources allocation algorithm which gives predictions and discounted rates in the tariffs by ensuring amount of resources reserved in the cloud.

2 RELATED WORK

The web based applications has been widely studied in the literature for prediction of CPU utilization and user access demand.The prediction method based on Radial basis function(RBF)networks has been proposed by Y.Lee.et.al to predict the user access demand request for web type of services in web based applications.Prediction method is to improve the capability for making informed decisions by providing reliable functions based on neural networking and linear regrecision.This were Radial basis functions which is proposed by Y.Lee.et.al for predicting the user access demand request for web types of services in web based applications and output of network is a linear combination of Radial basis function of inputs.Web application and demand prediction for CPU utilization has been studied for long time to gain the popularity recently. Cloud providers offers a

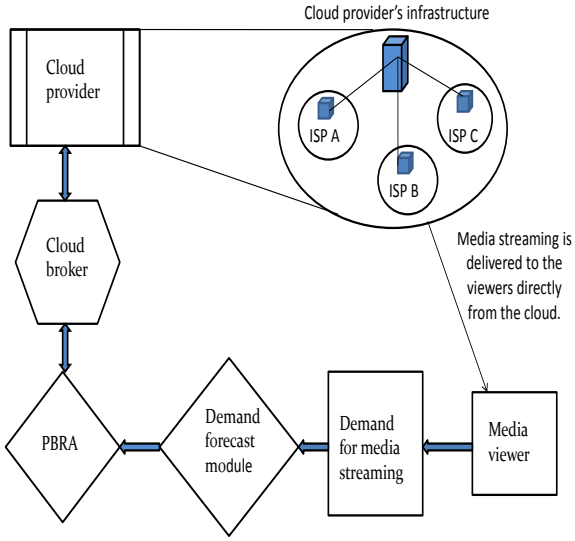
streaming resources to content providers with intensive bandwidth demand.

3 PROPOSED SYSTEM

Here we implement an innovative algorithm based on prediction based resource allocation algorithm in which resource allocation is the process and strategy involving a company deciding where scarce resources should be used in the production of goods and services. BRA gives predictions and minimizes the cost of resources in the cloud by giving discounted rates in the tariffs. Here predictions are made in such a way that it considered decision from cloud service provider. The main view of cloud service provider is it should optimize the memory resource allocation so that it can provide to other content provider. Cloud service provider provides two services to customers they are linear tariff plans and non-linear tariff plans. The main aim is to reduce our expenses by choosing a right plan at right time by ensuring sufficient resources which considerably gives more profit.

The content provider reserves resources in the cloud according to the predicted demand. The proposed algorithm is based on time-slots with varied durations (sizes). In every time-slot, the media content provider makes a decision to reserve amount of resources in the cloud. Both the amount of resources to be reserved and the period of time over which the reservation is made (duration of time-slots) vary from one time-slot to another, and are determined in our algorithm to yield the minimum overall monetary cost (Fig. 2). We alternatively call a time-slot a window, and denote the window size (duration of the time-slot) by w. Since the actual demand varies during a window size, while allocating the resource in the cloud remain the same for the entire window size (according to the third assumption above), the algorithm needs to reserve resources in every window j that are sufficient to handle the maximum predicted demand for streaming capacity during that window with some probabilistic level of confidence η.

We denote the monetary cost of the reserved resources during window j by Cost (wj, Allocj), and can be computed as Cost (wj, Allocj) = tariff (wj, Allocj)×wj, (2) where tariff(wj,Allocj) represents the price (in $ per time unit) charged by the cloud provider for amount of resources Allocj reserved for period of time (window size) wj. More specifically, the demand forecast module predicts streaming capacity demand in the upcoming period of time L and feeds this information to our algorithm .The algorithm upon receiving the demand prediction, computes the right size of window j (i.e., w∗ j), and the right amount of reserved resources in window j (i.e., Alloc∗ j), , such that the cost of the reserved resources during window j (i.e., Cost(wj,Allocj) in (2)) is minimized; or equivalently, the discounted rates offered in the tariffs are maximally utilized. Hence, the objective of our algorithm is to minimize Cost(wj,Allocj) ∀j, subject to Probability(D(t) ≤ Alloc(t)) ≥ η, ∀ t ∈ L. In other words, our objective is to minimize the monetary cost of reserved resources such that the amount of reserved resources at any instant of time is guaranteed to meet the actual demand with probabilistic confidence equals to η.



SYSTEM DESIGN

Fig1 :System Architecture

Table 1. Summary of results for iterations executed for window j = 1

iteration(h)	1	2	3	4	5	6
wh	1.0	2.0	3.0	4.0	5.0	6.0
_max	0.6	1.0	1.184	1.26	1.3	1.37
Alloch	3	1.0	2.0	2.0	2.0	2.0
Xh=tariff(wh,Alloch)	1.0 11	10.85	12.25	12.0	11.75	11.50

Table 2.Media streaming cost given different probability of the demand (in $)

Distribution	Long-normal(σ=3)	Long-normal(σ=6)	Long-normal(σ=8)
cost	34,457	41,543	48,393

Table 3.Media streaming cost using two resource allocation plans provided by the cloud (hybrid resource provisioning approach) (in $)

n	Cost of reservation plan	Cost 0f on demand plan	Total cost
0.75	34,457	12,213	46.670
0.8	36,979	8,854	45,833
0.9	44,033	2,821	46,854
0.95	46,324	2,741	49,065

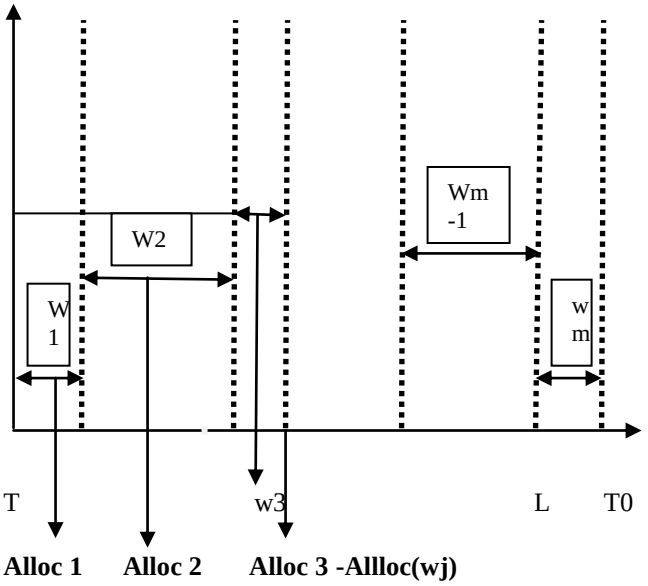


Fig2: PBRA algorithm design

wj: the jth window size M: number of windows Allocj: amount of allocated resources in window j

4 RESULTS AND DISCUSSION

To reduce our expenses by choosing a right plan at right time by ensuring sufficient resources in the cloud which comfily gives more profit? We have considered Non-Linear tariff plans because it is less costlier, more flexible, gives more discounts from customer perspective and non-fixed with respect to customer's perspective. The proposed algorithm based on time slots with varied durations. In every time slot the media content provider makes a decision to reserve amount of resource in the cloud. Both the amount of resources to be reserved and period of time over which reservation is made vary from one time slot to another and it is determined in our algorithm to yield the minimum overall monetary cost. Media content provider can predict the demand for streaming capacity of a video channel over a future period of time and it reserves resources in the cloud according to the predicted demand.The proposed algorithm PBRA gives predictions and minimizes the cost of resources in the cloud by giving discounted rates in the tariffs and PBRA provides demand to avoid unnecesaary amount to be paid to eliminate the wastage. The results show that our algorithm adjust the trade-off between the resources in the cloud and support well defined plan such that as much as the demand is coming the trade must be optimized.

Fig3:Cloud users usage profile

Fig4:Cloud users validity profile

Fig5:Cloud users on demand video usage

Fig6: Cloud users increment over the years

5 CONCLUSION

This paper studies the problem of resource allocations in the cloud for media streaming applications. We have considered non-linear time-discount tariffs that a cloud provider charges for resources reserved in the cloud. We have proposed algorithms that optimally determine both the amount of reserved resources in the cloud and their reservation time - based on prediction of future demand for streaming capacity - such that the financial cost on the media content provides minimized. The proposed algorithms exploit the time discounted rates in the tariffs, while ensuring that sufficient resources are reserved in the cloud without incurring wastage. We have evaluated the performance of our algorithms numerically and using simulations. The results show that our algorithms adjust the trade-off between resources reserved on the cloud and resources allocated on-demand. In future work, we shall perform experimental measurements to characterize the streaming demand in the Internet and develop our own demand forecasting module. We shall also investigate the case of multiple cloud providers and consider the market competition when allocating resources in the cloud.

6 REFERENCES

[1] Cisco Systems Inc., "Cisco visual networking index: Forecast and methodology, 2010-2015," white paper, 2010.
[2] Y. Liu, Y. Guo, and C. Liang, "A survey on peer-to-peer video streaming systems," in Peer-to-Peer Networking and Applications,vol. 18, no. 1, pp. 18–28, 2008.
[3] Cisco Systems Inc., "Data center virtualization and orchestration: Business and financial justification," white paper, 2007.
[4] Four Reasons We Choose Amazons Cloud as Our Computing Platform, The Netflix Tech Blog, Dec., 2010.
[5] http://www.octoshape.com/
[6] Amazon EC2 Reserved Instances, http://aws.amazon.com/ec2/reserved-instances, 2012.
[7]AmazonCloudFront, http://aws.amazon.com/cloudfront/, 2012.
[8] G. Chuanxiong, G. Lu, H. Wang, S. Yang, C. Kong, P. Sun,H. Ballani, P. Costa, T. Karagiannis, and A. Rowstron, "Towards predictable datacenter networks," in Proc. of the ACM SIGCOMM conference, pp. 242–253, 2011.
[9] E. White, M. O'Gara, P. Romanski, P. Whitney, "Cloud Pricing Models," in Cloud Expo: Article, white paper, 2012. http://java.sys-con.com/node/2409759?page=0,1.

An overview of Swarm Intelligence based Algorithm for Optimization Problem in Wireless Sensor Networks

[1]*Nandini G,* [2]*Dr J Anitha*
[1]*Research scholar, VTU-RRC,* [2] *Professor, Dept of CSE, DSATM, Bengaluru.*
nanduamma@gmail.com, anitha.jayapalan@gmail.com

ABSTRACT- Wireless sensor network consist of many sensor nodes in which each sensor node collects the data from sensing environment and transmit it to the Ebase station. The most important key in WSN's is to find the optimal path to transmit the data. Nature is the best coach and its designs and strengths are enormously massive and eccentric that it gives inspiration to researchers to emulate nature to solve hard and complex problems in networks. Computation using Bio inspired algorithm has come up as a new epoch in computing, casing wide range of application. This paper gives overview of most successful classes of swarm intelligence (SI) based algorithm for solving optimization problem.

Keywords: wireless sensor networks, optimization problem, swarm intelligent based algorithm.

I. INTRODUCTION

Wireless sensor network is a group of large number of sensor nodes that are deployed in a region which manage themselves to form a network and work collaboratively to perform network functionality. WSN's can be used in many applications like healthcare, military, battle field and tracking and monitoring systems. For example sensor networks can be deployed in the atmosphere for monitoring and controlling of plant and animal behavior or in ocean for controlling temperature.

Fig 1: Simple Architecture of Wireless Sensor Network.

There are some explicit characteristics of WSNs where routing protocols must have in order to make use in real-world applications.
1) Memory requirements and minimal computation.
2) Self-organization.
3) Energy efficiency.
4) Scalability.
5) Support for in-network data aggregation.
Swarm intelligent algorithm are the problem solving methodology which is derived from the natural behavior and operation of natural system. These algorithms can solve problems in almost all application areas like wireless sensor networks, computer networks, image processing, mining data, signal processing, control systems and more. Bio inspired algorithm have the ability to resolve the complex relationship from inherently very simple initial conditions and rule. The architecture of wireless Sensor Networks is a s shown in the Fig 1.

The organization of paper as: section II provides the need for optimization in WSN's. Section III provides motivation behind swarm intelligent family. Section IV provides overview of various swarm intelligence family. Conclusion is drawn in section V.

II. NEED FOR THE OPTIMIZATION IN WSN'S

Optimization is the process or method for obtaining the most excellent results under a given condition. Network optimization is a very significant task and a number of optimization techniques are used to attain desire goals in network. The optimization of both hardware and software is required for designing of WSN's for achieving energy efficiency, reduced cost and application requirement optimization. For different problems several optimization algorithm are present, the challenge is that selecting a best algorithm for a requirement. More important requirements for a WSN are, achieve high quality QoS, energy efficiency low bandwidth, limited processing and storage in sensor node [6]. These are the some of the issues of WSN which are directly related to the problem of optimization.
Sensor networks and biological systems require to adapt themselves according to the changeable environmental conditions including the ability to selforganize, scalability and to provide robust operation for the long life of the sensor network. [7]. The simple optimization process is as shown in the Fig 2.

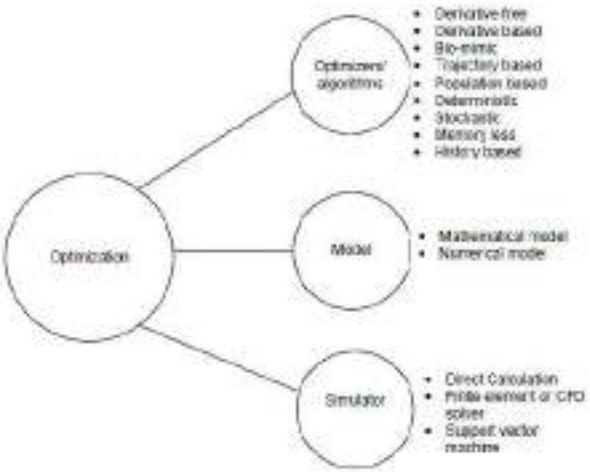


Fig 2 A simple optimization process.

III. MOTIVATION BEHIND SWARM INTELLIGENCE (SI) FAMILY

An intensive literature survey on nature inspired algorithm has been carried out to know the cause for intrinsic weakness of evolutionary optimization algorithm. The survey concealed that the performance of existing algorithm are dependent on time taken by biological counterpart. For example in genetic algorithm performance is slow in retrieving optimal solution as it implementation of simple Darwin theory which takes number of years to reveal changes in fitness of species. In Differential evolution according to krink et al. noise adversely affect the performance of Differential evolution because of its greedy nature. The best values for problem dependent parameters have to be found out by user, which is a time consuming task.

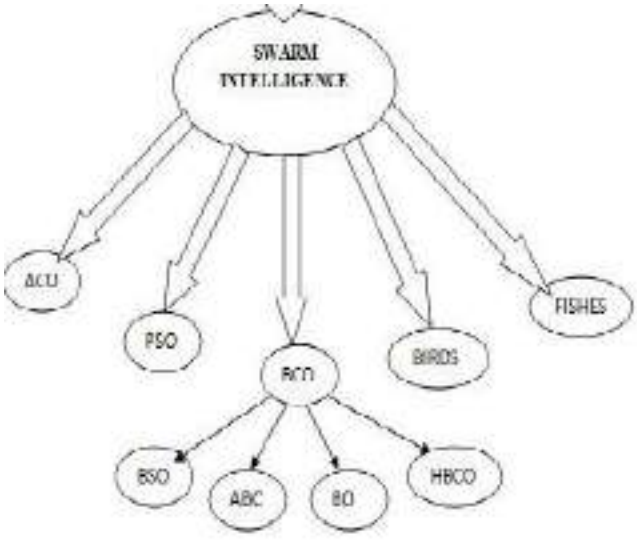


Fig 3: hierarchical representation of Swarm intelligence family.

IV. OVERVIEW OF SWARM INTELLIGENCE FAMILY

A Swarm can be defined as a set of mobile agents which are likely direct or indirectly communicate among each other and collaboratively solve a distributed problem. Swarm intelligence is the area which deals with nature and system collected individuals that synchronize using self organization and decentralization.

A very simple rules are followed by agents even though there is decentralized control structure which dictates agents how to behave, interaction between agents. Natural examples of SI consist of ant colonies, fish schooling and bee colony etc.

There are five fundamental principles to describe swarm intelligence.

1)principle of proximity: to carry out simple space and time computation.

2) principle of quality: responding to quality factors in the environment.

3)principle of diverse response: should not commit activity along excessively narrow channels.

4)principle of stability: with environment changes should not change its mode of behavior.

5)principle of adaptability: change behavior mode when worth of computational price.

The classification of swarm intelligence algorithm is on the basis of social behavior of animals, human immune system, among which algorithm based on behavior of animals can be explored as:

A Particle swarm optimization (PSO)

PSO algorithm is based on population based optimization technique proposed by Kennedy and Eberhart [9]. Due to its simple concept, efficient computation, easy implementation and unique searching mechanism PSO has been used in many engineering problem. The steps of PSO are as below:

1. Swarm has to be initialized by assigning a random position.
2. Fitness function has to be estimated for each particle.
3. Compare particle's fitness value with pbest for each individual particle. If the current value is better than the pbest value, then set this as pbest for current paticle's position.
4. Particle is identified which has best fitness value and is identified as gbest.
5. Velocities and positions of all the particles are revised using step 1 and 2.
6. Until sufficiently good fitness value is achieved repeat steps 2-5.

Advantages of PSO:

1. It is easy to implement and there are few parameters to adjust.
2. It has more effective memory capability.
3. In order to improve themselves PSO maintains diversity as all the particles use the information related to most successful particle.

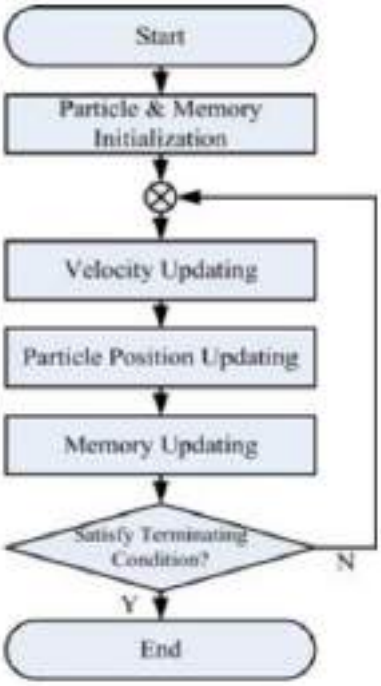


Fig 4: Steps for Particle swarm optimization.

B. Ant Colony Optimization (ACO)

in the nineties Dorigo et al. (1996) named Ant System‖ (AS), The first ant algorithm, tested successfully on the Travelling Salesman Problem. To generalize, the overall method of solving combinatorial problems by estimated solutions based on the generic behavior of natural ants, The ACO meta heuristic was developed (Dorigo & Di Caro, 1999;).the three main functions of ACO is as follows.

AntSolutionsConstruct: this performs the construction process of solution where according to a transition rule the artificial ants move through adjacent states of a problem, iteratively building solutions.

Pheromone Update: This performs updating of pheromone trail. updating the pheromone trails is done, once complete solutions have been built, or updating after each iteration. In addition to pheromone trail reinforcement, ACO also includes pheromone trail evaporation. Evaporation of the pheromone trials helps ants to forget‘ bad solutions that were learned early in the algorithm.

DeamonActions: is the optional step in the algorithm that involves applying extra updates from a global perspective. This may include applying supplementary pheromone reinforcement to the best solution generated. To improve the performance of ant system, Dorigo and Gambardella (1997) proposed an alternative approach, ant colony system (ACS), which is based on four modifications for ant system: a different transition rule, a different pheromone trail update rule, the use of local updates of pheromone trail to favor exploration, and the use of candidate list to restrict the choice.

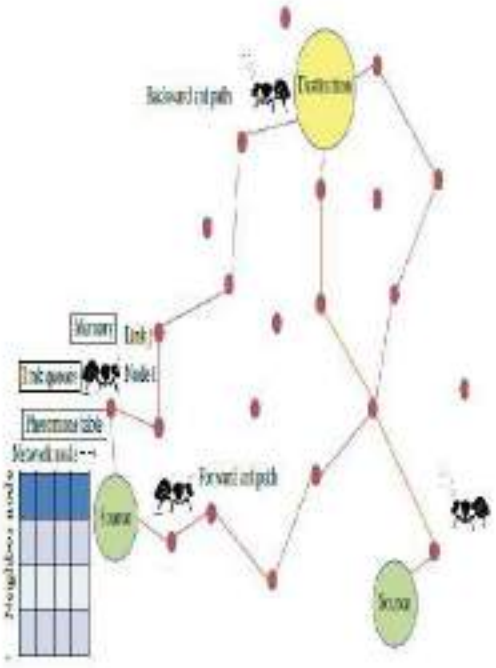


Fig.5. Representation of ACO Algorithm With Respect to Ants Foraging Behavior

C Bee Colony Optimization (BCO)

It is another Swarm Intelligence (SI) algorithm where the agents of the group are honey bees. Bees communicate with each other by a mechanism called “Waggle Dance”. They exchange information among themselves regarding the rich food source location. Because of their collective food foraging behavior, the name (BCO) is given to the honey Bees. The bee system is a standard example of organized team work, labor division, simultaneous task performance, specialized individuals. there are different types of honey bees In a honey bee colony. There is a queen bee, male drone bees and thousands of worker bees.

Types of bees: Queen: The responsibility of the Queen’s is to lay eggs to form new colonies . The responsibility of Drones which are the male of the hive is to mate with the Queen Bee .it will be discarded from the colony during their down fall. The worker bees are the females of the hive [19]. They are the important building blocks of the hive. They are responsible for building the honey bee comb, cleaning it, maintain it, guard it, and feed the queen and drones. Apart from these , the main job of the worker bee is to search and collect the food. Two types of worker bees are there namely scout bees and forager bees. Both of them are equally responsible for the collection of food but with different roles. The role of Scout Bee’s is to fly randomly all around and search for food. They come back to their hive after their exhaustion of energy and distance limits. Upon arrival to their hive the scout bee shares their

experience and a lot of important information with the forager bees.

D. Artificial Bee Colony Algorithm (ABC)

Various swarm intelligence algorithms are available Based on the behavior of the bees in nature, that are classified into two methods; foraging behavior and mating behavior. Algorithms simulating the foraging behavior of the bees include the Artificial Bee Colony (ABC) ,the Virtual Bee algorithm , the Bee Colony Optimization algorithm , the Bee Hive algorithm, the Bee Swarm Optimization algorithm. and the Bees algorithm. An individual entity demonstrate a simple set of behavior policies , but a group of entities shows complex emergent behavior with useful properties such as scalability and adaptability. Artificial Bee Colony is a predominant algorithm simulating the intelligent foraging behavior of a honeybee swarm, proposed by Karaboga and Basturk .In ABC algorithm, the colony of artificial bees contains three groups of bees: *employed bees, onlookers and scouts.*
An employed bee produces a modification on the position in the memory depending on the loca linformation and tests the nectar amount of the new source. Using the nectar amount of the new one is higher than the previous one, the bees remember the new one and forget the old one. After every working bees complete the search process. They share the nectar information of the food sources and their position information with the onlooker bees on the dance area. The algorithm global search performance depends on random search process performed by scouts and solution production mechanism performed by employee and onlooker bees.

E. Fish Swarm Algorithm (FSA)

The fish swarm algorithm (FSA) is a population-based/swarm intelligent technique that is inspired by the natural schooling behavior of fish.FSA exibits a strapping ability to avoid local minimums in order to achieve global optimization. In this algorithm fish is represented by its D-dimensional position Xi = (x1, x2, . . ., xk, . . ., xD), and FSi to represent the food satisfaction for the fish. Euclidean distance dij = ||Xi −Xj|| denotes the relationship between two fish. FSA typically imitates three behaviors, which is defined as ―searching food, ―swarming in response to a threat, and ―following to boost the chance of achieving a flourishing result.
Searching can be said as as a random search adopted by fish for searching of food, with a tendency towards food concentration. The main objective is to minimize FS (food satisfaction).
Swarming: this aims in satisfying food ingestion needs, entertaining and attracting new swarm members. A fish which is located at Xi and has its neighbors within its visual . where Xc identifies the center position of those neighbors and it is used to describe the attributes of the entire neighboring swarm. If the swarm center has greater attentiveness of food than is available at the fish's current position Xi (i.e., FSc < FSi), and if the swarm (Xc) is not very crowded (ns/n < δ), the fish will move from Xi to next Xi+1, toward Xc .
Following behavior is one, when a fish locates food, all neighboring individuals follow it. Within a fish's visual, few fish will be apparent as finding a greater amount of food than others, and this fish will obviously try to follow the best one(Xmin) in order to increase satisfaction(i.e., gain relatively more food[FSmin < FSi] and less crowding[nf/n < δ]). nf represents number of fish within the visual of Xmin. Three important **parameters** involved in FSA include visual distance (visual), maximum step length (step), and a crowd factor. These two factors influence the effectiveness of FSA.

F. Bat algorithm (BA)

In 2010 yang proposed Bat Algorithm which is also a swarm based metaheuristic algorithm inspired by a property called as echolocation. This echolocation is a type of sonar which guides bats during flying and hunting behavior. two components which affect the search characteristics of an algorithm are exploration and exploitation.

Exploration is nothing but a capability of an algorithm to find solution which is promising by seeking strange region,
Whereas the exploitation improves the solution obtained by exploration. In survey many studies indicates that exploration capability of an algorithm must be employed first, and from which algorithm scans the whole search space and to improve the solution obtained by exploration should be employed in the optimization process.

G. Cuckoo search (CS)

CS is an optimization algorithm which was developedby Xin-She Yang and Suash Deb in the year 2009. This CS algorithm was inspired by the obligate brood parasitation of cuckoo species which lays their eggs in the nests of other host birds.

Some host birds engages direct conflict with the intruding cuckoos. For example, if the host birds discovers that the eggs in their nests are not their own eggs,then they will either throw the alien eggs away or abandon their nets and just build another one. The female parasitic cuckoos of New World brood-parasitic Tapera are very specialized in the mimicry in colors and pattern of eggs which resembles the eggs of the chosen host species.

Cuckoo search idealized such breeding behavior andthought of using this idea for various optimization problems. This idea can outperform other metaheuristic algorithm which are in applications.

Cuckoo search uses the following representation

1. Each egg in a nest is solution, and a cuckoo egg is a new solution .the aim of CS is to replace the not so good solutions in the nests with the new solutions(cuckoo). Now, this algorithm is like each nest has one egg.but this algorithm can be extended such that each nest has multiple eggs which means multiplication.
2. The idealization rule of CS is based on,

Each cuckoo lays one egg at a time, and dumps in random nests.

3.The best nests with high quality egg will be carried to next generation

The egg laid by a cuckoo is discovered by the host bird with a probability of a p_a\in(0,1). Discovering operate on some set of worst nests,and discovered solutions is dumped from farther calculations

CONCLUSION

All SI-based algorithms make use of multi-agents, which is inspired by the collective behaviour of social insects, like ants, bees, as well as from extra animal societies like flocks of birds or fish. The particle swarm optimization (PSO) uses the swarming behaviour of fish and birds, Ant colony optimization (ACO) uses the communication of social insects (e.g., ants), while the ABC Bee algorithms are all based on the foraging behavior of honey bees. Cuckoo search (CS) is based on the way of some cuckoo species brooding parasitism. Bat algorithm makes use of echolocation of few foraging bats. SI-based algorithms are the most popular and widely used algorithm for solving optimization problem in wireless sensor networks. There are many reasons for this, first reasons is that SI-based algorithms will share information among multiple agents, so that self-organization, co-evolution and learning through iterations will help to provide the high efficiency of most SI-based algorithms. Second reason is that multiple agent can be parallelized effortlessly so that significant optimization becomes more practical from the implementation point of view.

REFERENCES

[1] Jennifer Yick, Biswanth Mukherjee, Dipak Ghosal, "Wireless Sensor network survey", International Journal of Computer networks, Vol. 52, pp. 2292- 2330, 2008.

[2] C. Raghavendra, K. Krishna, T. Znati, "Wireless Sensor Networks", Springer-Verlag, 2004.

[3] Kemal Akkaya, Mohamed Younis "A survey on routing protocols for wireless sensor networks", IEEE Communication Magazine on Ad Hoc Networks, pp.325–349, 2005.

[4] Beyer, H.G. and Schwefel, H.P. 2002: Evolution strategies. Natural Computing 1,3–52.

[5] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, Journal of Global Optimization 11 (1997) 341–359.

[6] Upeka Premaratne , Jagath Samarabandu, and Tarlochan Sidhu, ―A New Biologically Inspired Optimization Algorithm‖,Fourth International Conference on Industrial and Information Systems, ICIIS 2009,28-31 December 2009, Sri Lanka.

[7] Bonabeau, E., Dorigo, M. and Theraulaz, G.1999: Swarm intelligence. Oxford University Press

[8] Kennedy, J.; Eberhart, R. (1995). "Particle Swarm Optimization". Proceedings of IEEE International Conference on Neural Networks. IV. pp. 1942–1948.

[9] Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant System: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics – Part B, 26, 29–41.

[10] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of Global Optimization 39 (2007) 459–471

[11] X. Li, Z. Shao, J. Qian,Anoptimizing method base on autonomous animates: fish- swarm algorithm, Systems Engineering Theory and Practice 22 (2002) 32–38.

[12] Dressler, F, Akan, O.B, "A survey on bio-inspired networking",
Computer Network, Vol. 54, pp.881–900, 2010.

[13] Clerc, M. and Kennedy, J. "The particle swarm-explosion, stability,
and convergence in a multidimensional complex space", IEEE transactions on Evolutionary Computation, Vol. 6, No. 1, pp. 58- 73. 2002.

[14] Hongliang Ren, Max Q H Meng, "Bio Inspired approches for Wireless Sensor Networks", IEEE Conference on Mechatronics and Automation A Survey", pp. 762-768, 2014.

[15] Bonabeau, E., Dorigo, M. and Theraulaz, "Swarm intelligence",
Oxford University Press, 1999.

[16] R. Storn, K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces", Journal of Global Optimization, Vol. 11, pp. 341–359, 1997.

[17] Dorigo, M., Maniezzo, & Colorni, "Ant System: Optimization by a colony of cooperating agents", IEEE Transactions on Systems, Vol. 26, pp. 29–41, 1996.

[18] A. Kaur and S. Goyal, "A Bee Colony Optimization Algorithm for Fault Coverage Based Regression Test Suite Prioritization", International Journal of Advanced Science and Technology, Vol. 29, pp. 17-30, 2011.

[19] P.Navrat, T. Jelinek, and L. Jastrzembska, "Bee hive at work: A problem solving, optimizing mechanism," World Congress on Nature & Biologically Inspired Computing, pp. 122 - 127 , 2009.
http://doi.acm.org/10.1145/1883612.1883618

[20] J. Kennedy, R. Eberhart, "Particle swarm optimization," Networks, 1995, IEEE International Conference on Neural Networks
, vol.4, no., pp.1942-1948 vol.4, Nov/Dec 1995

[21] R. Poli, J. Kennedy and T. Blackwell, " swarm optimization: An Overview", Swarm Intelligence, pp. 33 Springer, New York, 2007.

[22] M. Dorigo and T. Stutzle, "Ant Colony Optimzation", The MIT Press, Cambridge, Massachusetts, 2004
http://doi.acm.org/10.1145/1883612.1883618

A Scheduled Based MAC Protocols for Wireless Sensor Network: A Survey

Anitha K Research Scholar, **Usha S** Prof. & HOD
Dept. of Computer Science & Engg, RRCE, Bangalore
.anithakrishna14@gmail.com, sakthivelusha@gmail.com

ABSTRACT-The wireless sensor network has been gaining popularity in all fields. The applications of the WSN are traffic analysis, environmental monitoring, tactical systems, etc. Drawbacks of the WSN include limited battery power, limited bandwidth, memory constraint, etc. Hence designing an efficient MAC layer protocol for WSN is a challenging task. TDMA based MAC protocols can overcome the mentioned drawbacks. It can avoid collisions, idle listening and therefore is energy efficient. In this paper we have represented merits and demerits of different TDMA based MAC protocols for WSN.

Keywords - Centralized TDMA , Distributed TDMA ,Time Division Multiple Access(TDMA), Wireless Sensor Network(WSN),

I. INTRODUCTION

Recent advancement in the technology has enabled low cost, low weight and more energy efficient sensor devices. These sensor devices include integrated sensing, onboard processing and transmission. These sensor devices also known as sensor node is capable of detecting environmental conditions and sound. Sensor nodes have limited battery power leading to less coverage WSN consists of large number of sensor nodes to perform a designated task. In most wireless networks, collision [1] is caused due to two nodes sending data at the same time on the same transmission medium. To overcome this problem, MAC protocols are employed to the sensor networks. TDMA based MAC protocols time is divided into time frames. Each time frame is divided into fixed time slots. Each node is assigned to particular time slot in which transmission occurs. TDMA based MAC protocol [2] avoids the over hearing and idle listening since energy node has a fixed time slot for transmitting and receiving. Every node after receiving or transmission of data enters sleep mode i.e., switch off, thereby saving battery power. The rest of the paper organized as follows: Section II discusses about classification of MAC protocols followed by Section III describes the Scheduled based MAC Protocols and Section IV concludes the paper.

II. CLASSIFICATION OF WSN MAC PROTOCOLS

In WSN MAC protocol is categorized into Contention Based MAC protocols and Scheduled Based MAC Protocols .In Contention Based MAC Protocol, all the nodes share a common medium and contend for the same medium for transmission. Thus, collision may occur. To avoid collision, nodes can have arbitrary access to the shared channel. The sender listens to the shared medium before the transmission, waits for some time if the medium is busy, then tries for the transmission again. Examples of contention based MAC protocols are: Carrier Sensor Multiple Access (CSMA), Multiple accesses with collision avoidance, (ALOHA).

In Scheduled Based MAC Protocols, nodes access to the shared medium is divided with respect to time (Time Division Multiple Access), frequency (Frequency Division Multiple Access) and orthogonal pseudo codes (Code Division Multiple Access).This allows the different nodes to access the shared channel without collision. These protocols consume less energy hence they do not waste energy in collision and idle listening. Examples of Scheduled based MAC protocols are: Low energy adaptive clustering hierarchy (LEACH), Power efficient and delay-aware medium access protocol (PEDAMACS) etc.

Features of good MAC protocol

- Energy efficiency: The sensor nodes are battery charged and it has to be recharged frequently. Sometimes it's better to replace the sensor nodes rather than recharging.
- Latency: It refers to the time delay between time when data is sent by the sender and the time when data is received by the receiver. It depends on the application the detected quarts must be reported, so that the designed table is achieved.
- Throughput: It refers to the amount of data successfully transferred from a sender to the receiver in the given time similar to destiny. Its requirement depends on required application.
- Fairness: When bandwidth is limited, it is required to ensure that the sink nodes receive information from all sensor node fairly.

Among all the above features energy efficiency and throughput are the major aspects.

III. SCHEDULED BASED MAC PROTOCOLS

Scheduled based MAC protocol is divided into two categories. The first one is traditional TDMA MAC protocol or centralized TDMA protocols. Second category is distributed TDMA protocols.

In centralized TDMA protocols; the nodes are scheduled centrally by different time slots. The scheduling is done by cluster head (CH) or the base station (BS). In this method

if any mode join or leave the network then once again cluster head need to re-assign the scheduling time slots for each node. Disadvantage of this method is scalability and global time synchronization is required for all the nodes in a network.

A Centralized TDMA protocols

1) Bit-map resisted MAC protocol

Bit-map resisted MAC Protocol [3] is an Intra-cluster communication MAC protocol for a large scale cluster-based WSNS.BMA operates in two rounds. Fig 1 shows a complete round. Each round is divided into set-up phase and steady state phase. During set-up phase cluster-head is elected and determined based on the energy levels. The steady state phase is divided into n-sessions with fixed time duration. Each session consist of contention period, in this period all nodes keep their radio on and cluster-head has complete information about the nodes in the network. In TDMA period, nodes send a 1 bit control message to the cluster-head. When it has data to transmit, otherwise, its scheduled slots remain empty.

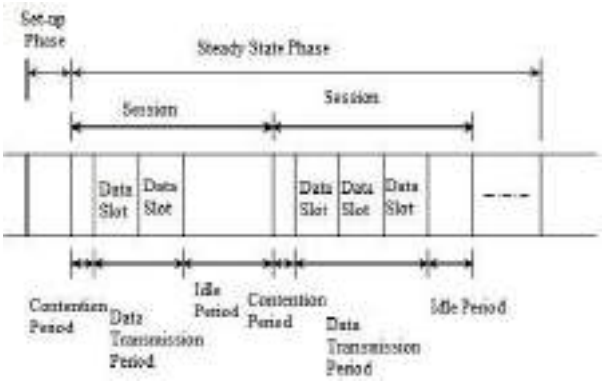


Fig1Single Round in a BMA

The cluster head decides the slot and broadcast to the source nodes to transmit data. During contention period slot is assigned to source node. In data transmission period, the node sends a data to cluster head during their allocated time slot and all other time node switch off the radio. When a session finishes, the next session begins with a again contention period followed by data transmission period. The same procedure repeated. The cluster head collects the data from all the node and forwards to base station.

Advantages of this algorithm is that it effectively utilizes the bandwidth, and average packet latency. Disadvantage of this algorithm is that it gives better result in low and medium traffic not for heavy traffic.

2) Self-Organized TDMA Protocol for WSN

Self-Organized TDMA Protocol [4] is a cross layer protocol to serve the application specific and data centric nature of WSN. In SOTP transmission range is very high. In this algorithm time slots divided into frames and each frame divided into time slots. Each frame consists of five types of time slots: Broadcast slot (BR), Carrier Sensing Slot (CS), Transmitting Slot (TX), Receiving Slot(RX) and Idle Slot(ID).A BR slot is always the first slot in the frame, CS is the second slot. Each node move into a one of three states: Searching synchronized or registered. When node boots up is called as node in searching state. The base station periodically broadcast the time slot, then node picks any of the broadcasted time slot then it is called node moved to synchronized state. This picked time slot need to be informed to father node then node moved to register state. In registered state can have one father node and several child nodes.

The merit of SOTP is more energy efficient due to its pure TDMA and non-clustering architecture. The main drawback of this method is that it assumes the high transmission range it covers all the nodes in the network. Data aggregation and compression is given to the upper layer.

3) Mobility tolerant TDMA based MAC protocol

Mobility tolerant TDMA based MAC protocol is used in mobile wireless sensor network. It is assured that network is static during its set-up phase. The network is divided into different clusters and each cluster will have a cluster head. The time is divided into frames and in turn frame is divided into time slots as shown in Fig (2).

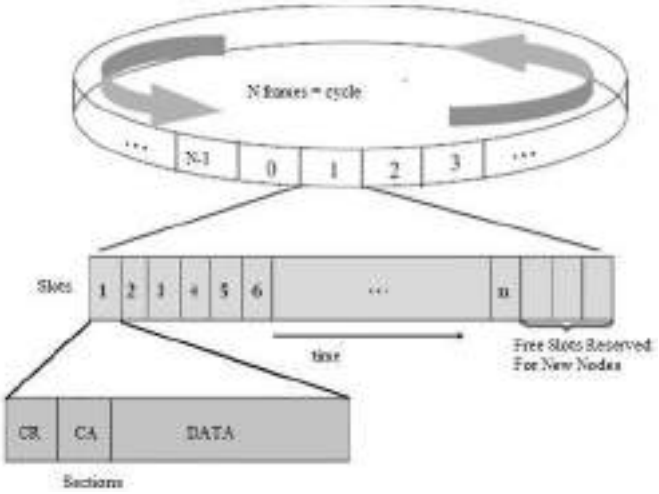


Fig2 Slot representation in a Mobility tolerant TDMA MAC

The cluster head allocated a time slot to every node in the cluster and X slots are made free in order to allocate for the nodes which join later part of the time frame. The 10% of total time frame is also allocated for Y. Whenever new node comes and joins the cluster, the cluster head allocates half of the free time slots (Y/2) to the new node. If another node joins to the network then cluster head allocates half of the current time slots (Y/4). Always cluster will a some free time slots to allocate to a new node that join the network.

In this method each time slot is further divided in to 3 sections: Communication request (CR), channel allocation (CA), and data section (DS). In communication request all those nodes that have data to transmit will put the request to the owner node. The owner node calculates the priority of each node. Based on the priority it allocates a time slot. In channel allocation section the packets are transmitted or received in data section. The node which is having highest priority, then owner node assigns first slot to highest priority node.

B Distibuted TDMA Protocols

In Distributed TDMA Protocols, nodes are scheduled by themselves. Each individual node will have the information based on these local information scheduling. The message overhead can be reduced by this method. So this method is more energy efficient. Global time synchronization is not needed. Distributed TDMA protocols are adaptive to topology changes and bandwidth adjustment. In the next subsection some of the Distributed TDMA Protocols are discussed.

1) Distributed Neighborhood Information Based TDMA Scheduling (DNIB)

The DNIB algorithm, composed of the following stages: Slot assignment, Update and Recovery

At initial stage nonscheduled node computes for contender rank for itself and its nonscheduled neighbors, contender rank are computed based on hop distance and the node id. A node itself as a rank 1, when it has a highest C value. The contender rank is updated each time a neighbor is scheduled. Node through the Update procedure informs to its neighbors. It sends updated message to its one-hop neighbors. Those one-hop neighbors send broadcast message to their neighbor's .Now the message is reached to two hop neighbors to update. The broadcast message contains node id and assigned slot. For two hop broadcast message it give slot assignment information for the one hop to reach to two hop neighbors. During this procedure, f collision may occur. The collisions between two-hop broadcast and one-hop broadcast messages may exist.so delay can be introduced for one hop message otherwise wait for a predetermined time slot ω. An example is illustrated in Fig 3.

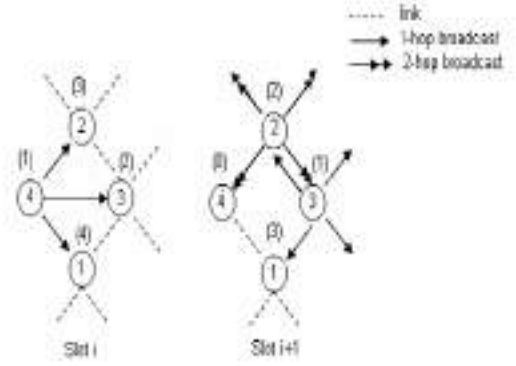


Fig3: Collision during updated procedure

The third procedure is Recovery procedure. This procedure is used when a node does not succeed to get scheduled for a predetermined period of time denoted Xn for node n. In this case, the node can send a "request" message to its one hop neighbors. This request message is the same as that of one-hop broadcast message except the assigned slot values which is set to NULL. Node sends a message to one-hop neighbors using "two hop broadcast" message with all already known scheduling information. If two nodes sending message at same time can be avoided by setting Xn.

The advantage of this algorithm is that, it is scalable. It can run parallelly in a network. It does not require whole network information, using a two hop neighbor it can perform the operation. The main demerit of this algorithm is recovery procedure is based on setting of Xn, if too small or too large time will be the problem for collision.

2) Distributed Randomized TDMA Scheduling (DRAND)

The DRAND is based on coloring problem. In order to assign TDMA slot assignment it uses graph coloring, where the no two adjacent nodes have the same color and vertices should use minimum color. In DRAND each node will be in a four states: Idle, Request, Grant, and Release.

Initially Node A will be in an Idle state. DRAND sends a request message based on probability, when we are tossing a coin probability of getting head or tail is ½ .The based on probability P Node A sends a request message to node B as shown in Fig (4).Node B adds node A time slot into the list.

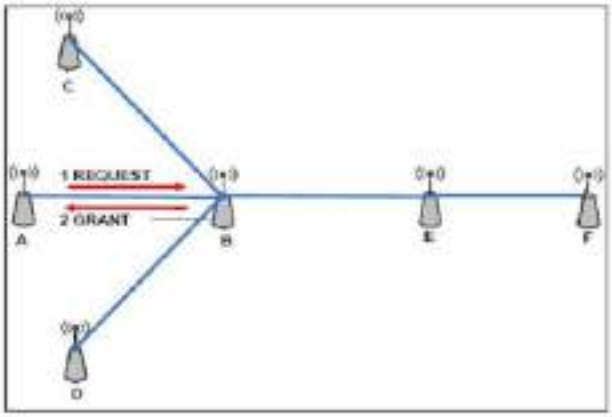


Fig4: Sucessful slot assignment by DRAND

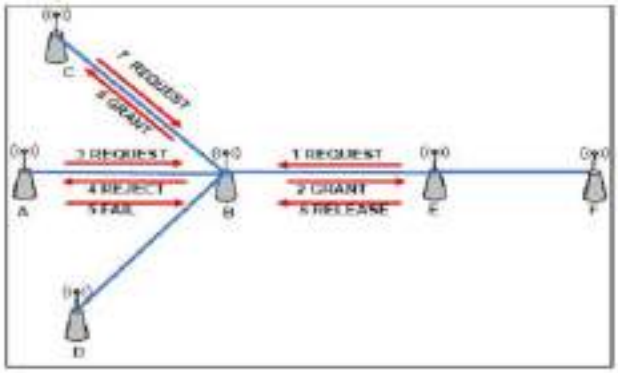


Fig5 Unsuccessful slot assignment by DRAND

If node B does not get a request message from any of the one-hop neighbor before node A, then node B gives a Grant message. In turn it sends a request message to two-hop neighbors. Based on node B grant message node A allocates a time slot .Node A will receive a Grant message only on node B in Idle or Release state. If node B receives Request message from one of its neighbor node D before getting a request from node A as shown in Fig (5). Nodes B send a Release message to node A, because it has already sent a grant message to some other node. In both the cases node A does

not receive Grant Message. Node A has to wait for a time period Ta (time delay node A receives response from neighbors).The node A sends a Fail broadcast message to all its neighbors and goto Idle state. After Ta period node A once again send request message.

An advantage of DRAND, scalable, slot assignment is calculated in each node locally. In large network it works efficiently. The main disadvantages of DRAND are message overhead. The energy is wasted for unnecessary message transmission.

1) Fair Scheduling TDMA Protocols

Fair Scheduling TDMA Protocols is a distributed slot assignment algorithm. Each Node maintains four states as that of DRAND IDLE, REQUEST, GRANT and RELEASE. Node E sends a request message to node B. If node B does not get a request message from any of the one-hop neighbor before node E, then node B gives a Grant message. In turn it sends a request message to two-hop neighbors. Based on node B grant message node E allocates a time slot .Node E will receive a Grant message only on node B in Idle or Release state. If node B receives Request message from one of its neighbor node C before getting a request from node E as shown in Fig (6). Node B store a request message in its queue and will send a WAIT message to the nodes which have sent REQUEST message as shown in Fig.6.Node B aware of REQUEST from all nodes. Node B sends a GRANT message based on priority, which node request first they will get first grant message.

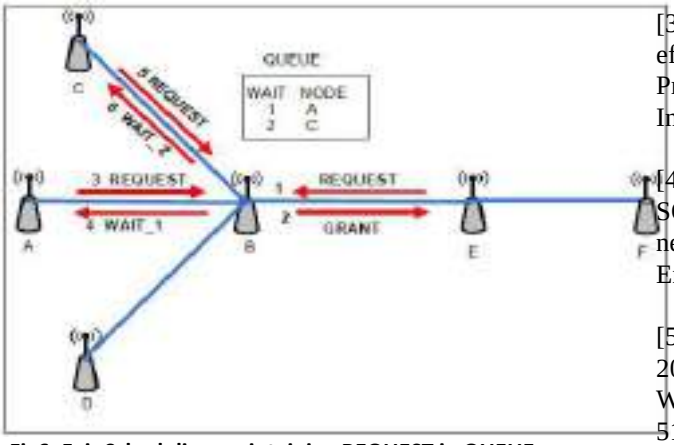


Fig6: Fair Scheduling maintaining REQUEST in QUEUE

The advantages of this method is that, message overhead can be reduced, more energy efficient. The main disadvantage is request message has to be stored in a node memory.

for nurses. They are used for recording home visits for the elderly. NFC is becoming widely accepted for medical devices in some markets specifically in the developed countries. Sony Corporation has developed an NFC Healthcare Library which enables communication between healthcare products embedded with the NFC Dynamic Tag (FeliCa Plug) and healthcare applications installed on smart phones. This library is available free of charge for a number of OS, including Windows, Linux and Android. Companies like Omron, Terumo and A&D are incorporating Sony's solution into their devices like BP monitors, pedometers, blood glucose meters, etc. Various other companies like Qolpac and Identive WPG have brought NFC into the mainstream with uses ranging from medication compliance to X-ray image sharing.

IV. CONCLUSION

In this Paper scheduled based MAC protocols are discussed. In all the scheduled based Protocols advantages and disadvantages are discussed, but still there is no standard protocol is defined .The scheduled based protocol is categorized into centralized and distributed. In centralized method cluster head or base station has to allocate a schedule, so cluster head or base station will have more overheads. In distributed approach each node in a network is responsible for allocating a time slot schedule. Compare to centralized one distributed gives better performance, but still some of the drawbacks were observed. Improvement need to be made in scheduled based protocols.

REFERENCES

[1]Chatterjee, P. and Das, N. 2009. A Cross-Layer Distributed TDMA Scheduling for Data Gathering with Minimum Latency in Wireless Sensor Networks‖, IEEE Wireless VITAE'09, 813.

[2] S. S. Kulkarni, ―TDMA Services for Sensor Networks,‖
Proc. 24th nt'l. Conf. Distrib. Comp. Sys. Wksps., Mar.2004, pp. 604–09.

[3] Li, J., Lazarou, G.Y. 2004. A bit-map-assisted energy-efficient MAC scheme for wireless sensor networks, Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks, 55-60.

[4] Wang, Y., Henning, I., Li, X. and Hunter, D. 2006. SOTP: A self-organized tdma protocol for wireless sensor networks, IEEE conference on Electrical and Computer Engineering, 1108-1111.

[5] Thaskani, S., Vinod Kumar, K., and Rama Murthy, G. 2011. Mobility Tolerant TDMA Based MAC Protocol for WSN, IEEE Symposium on Computers & Informatics. 515-519.

[6] Ines Slama, Bharat Shrestha, Badii Jouaber, Djamal Zeghlache, and Tapio J. Erke, DNIB: distributed neighborhood information based Tdma scheduling for wireless sensor networks, IEEE,2008.

[7] I.Rhee, A.Warrier, J.Min, and L. Xu, "DRAND: Distributed randomized TDMA scheduling for wireless

adhocnetworks," in Proc. ACM MobiHoc, New York, 2006, pp. 190–201.

[8] Muhammad Aman Sheikh, Micheal Drieberg and Noohul Basheer Zain Ali,"Fair Scheduling Algorithm for Wireless
Sensor Networks", IEEE,2011.

[9] I Rhee, A. Warrier, M. Aia, and J. Min. Z-MAC: a hybrid MAC for wireless sensor networks. In SenSys'05: Proceedings of the 3rd international conference on Embedded networked sensor systems, pages 90–101, York, NY, USA, 2005. ACM Press.

[10] Kalidindi, R, Ray , L., Kannan, R. and Iyengar, S.2003. Distributed energy aware MAC layer protocol for wireless sensor networks , in Proceedings of 2003 International Conference on Wireless Networks (ICWN ' 03) , Las Vegas, Nevada.

[11] Rajendran, V., Obraczka, K. and Garcia - Luna - Aceves, J. J. 2003. Energy-efficient, collision-free medium access control for wireless sensor networks, in Proceedings of 1st ACM Conference on Embedded Networked Sensor Systems (SenSys ' 03) , Los Angeles, CA, 181 – 192 .

NFC-Based Secure Mobile Healthcare System

Madhura PM, Palash Jain, Harini Shankar
Department of Information Science, Visveswaraya Technical University, Bangalore-60
madhura.manjunath7@gmail.com, palashjain2801@gmail.com, harini_27oct1994@hotmail.com

ABSTRACT-Near Field Communication (NFC) is a wireless communication standard which enables two devices in a short range to establish a communication channel within a short period of time through radio waves in the 13.56 MHz frequency range. A secured and an efficient architecture for improving healthcare system, uses Android based mobile devices with NFC smartcard technology on tamper resistant Secure Element (SE) and a Health Secure service on a hybrid cloud for security and health record management. In this work, we propose NFC cards (MIFARE classic cards) that are equipped with internal memory, which provide the potential advantage of improving patient's identification by eliminating the paper based documentation work. Simple touch of NFC enabled mobile devices can benefit both the patient as well as the medical doctors by providing a robust and secure health flow. It can also provide portability of devices and usability for health management in emergency situation, overpopulated hospitals and remote locations.

I. INTRODUCTION

Robust healthcare is a requirement for both developed countries, where the cost of healthcare is high and security and privacy are critical issues and developing countries like India. The advancement of science and technology in the field of healthcare has improved the quality of people's lives. At the same time mobile phones are gradually adopted for solving some tough healthcare issues. The purpose of this project entitled NFC BASED SECURE MOBILE HEALTHCARE SYSTEM is to computerize the front office management of hospital to develop software which is user friendly, simple, fast, and cost effective using NFC's promising features. As NFC cards are equipped with internal memory, patients can have critical information stored on their NFC card for fast access in critical situations. This information can vary and can be configurable. Information such as blood type and allergies can be stored as critical fast access data. This paper proposes a novel usage of NFC enabled mobile devices to access secure external medical tags for identifying medical objects like medicines and patient Health cards. The Health card could be on an external tag or retained on the patient mobile device using NFC Peer to Peer(P2P) or card emulation modes. The business logic of using Health card on mobile devices can be beneficial to a medical professional since it can securely identify patients using simple portable mobile devices and also get a concise health report. Simplified workflows will result in faster and more efficient patient-doctor interaction.

II. NFC TECHNOLOGY

NFC is an upcoming wireless technology which provides simple interfaces for device to device communication as well as access to NFC, Radio Frequency Identification(RFID) and smartcard tags. NFC enabled mobile device can operate in three modes: i) Reader mode: in which device can read and write to NFC based passive tags. ii) Peer to Peer (P2P) mode in which NFC devices can interact and exchange information with each other iii) Card emulation mode: in which NFC device can operate as a contactless card. NFC tags are of different types and use NDEF (NFC Data Exchange Format) for storing and sending data. We utilize MIFARE Classic 1K tags, which employ a proprietary protocol compliant to parts of ISO/IEC 14443-3 Type A, and write raw data using NFC-A (ISO 14443-3A) properties for improved security. The MIFARE Classic IK tag offers 1024 bytes of data storage, split into 16 sectors. Each sector is protected by two different keys, called key A and key B for secure access. NFC enabled mobile devices have a secure element (SE) which is a secure microprocessor (a smart card chip) that includes a cryptographic processor to facilitate transaction with authentication and security, and provides secure memory for storing applications and credentials. It is Java Card 2.2.2 compliant. Java Card is a technology which enables Java based applets to run on smartcards with very limited memory and processing capabilities and provides data encapsulation, firewall and cryptography[1]. The smart card specification standards, ISO/IEC 7816 for contact and ISO/IEC 14443 for contactless, specify that communication between a host application and a smart card is done through Application Protocol Data Units (APDUs). It is compliant to work on Android 2.3.3 and above.

III. PROPOSED ARCHITECTURE

We have proposed an architecture for NFC based secure health care as illustrated in Fig. 1 for i) secure medical identifiers as in flow steps 1.1 to 1.5 and ii) Health card retaining EHR using Android mobile devices as in flow steps 2.1 to 2.5. We have proposed a secure healthcare service like Health Secure on a hybrid cloud to which all hospitals can subscribe. The Health Secure hybrid cloud provides service for maintaining Cryptographic servers for secure framework and Storage server to provide backup as well as space for extended EHR. $Mobile_{ADMIN}$ is a mobile device of an authorized medical admin.

$Mobile_{pAT}$ is the patient's mobile device with the Health card and $Mobile_{Doc}$ is the doctor's mobile device. Since a larger screen would be better suited to view and update the health records, $Mobile_{Doc}$ could either be an NFC enabled tablet, for portability, or a laptop with external smartcard reader. K_A and K_B are the read and write access keys respectively for a secure tag based on MIFARE Classic. For NFC P2P based and card emulation based Health cards, we use patient's and doctor's set of public and private keys, which are Kp_{UBPAT}, Kp_{RIPAT} and Kp_{UBDOC}, Kp_{RJDOC} respectively. A symmetrical shared key K_{SH} is used for encrypting data. Hospital administration has an application for securely reading/writing with a mobile device, $Mobile_{ADMIN}$, to

manage smartcard based tags and patient Healthcards. $Mobile_{ADMIN}$ can register with the proposed HealthSecure cloud service on a hybrid cloud, which can issue security keys for our architecture. The mobiles use SE and simple interfaces of NFC and Bluetooth for credential storage and communication. We discuss the architecture of the applications briefly and the details of the implementation in section IV.

A. NFC Tags Utilization for Secure Medical Object Identification

It is important to reduce errors in the hospital workflow using Reliable medical object identifiers, such as giving correct medicine to a patient. We propose architecture of an application for issuing secure identifiers to reduce the error

Fig.1: Architecture of NFC based mobile healthcare system

and also to prevent security attacks like modification, repudiation and masquerading. The secure NFC passive tags have been used for identifiers, specifically MIFARE Classic. Bluetooth Low Energy (BTLE) stickers have lately been used to identify objects. But since they require a dedicated battery to operate, NFC passive tags are cheaper for identifiers to be used in healthcare. Hence basic NFC-A interface can be used to access smartcards from a mobile device. A valid mobile reader must have security key K_R for read access and a valid writer must have security key K_W for update access. The tag is issued by a healthcare admin mobile device, $Mobile_{ADMIN}$, which has registered to a HealthSecure service. It retains security keys in its SE for issuing tags. To enhance security, the access keys of the tag could be updated on a periodic basis for retaining secure IDs on the medical objects. Fig. 1, steps 1.1 to 1.5, shows the workflow of secure tag identifiers in bold. Along with medical identification records, information related to timestamp can also be updated.

B. NFC Tags in e-Health Cards

The secure tags used for application in III-A, are used for a different application for storing EHR on Healthcard of a patient. This is similar to a smartcard based Healthcard. But here we suggest smartcards that can be securely and easily be accessed using mobile devices. The tag could retain patient identification information along with emergency information, insurance information and health records. The tag could be organized into different sections, each administered separately by different set of security access keys. Similar to the secure tag application, this card can be issued and updated by an authorized health admin mobile device $Mobile_{ADMIN}$. A patient can register at the $Mobile_{ADMIN}$ and then later show to an authorized doctor with $Mobile_{DOC}$ in an OPD which would have the required access keys K_R and K_W for reading and updating the health records respectively. All NFC information can be retained with a timestamp. Detailed health records can be retained on a storage server of the HealthSecure service on hybrid cloud. At the end of the visit the patient can present the tag back to the administrator to tap and store his visit detail on the hybrid cloud. At any point of time if patients' past records are required, they can be retrieved over secure wireless interface (like HTTPS) from the hybrid cloud, using the patient ID on the tag. This application will help the patient to retain the recent health records on a cheap yet secure tag equivalent to a smartcard.

C. e-Health Card based on P2P NFC mode

This application architecture is based on retaining a Healthcard on a mobile device using NFC P2P mode. The EHR is retained on the mobile device in a secure region instead of NFC tag as in III-B. The patient can tap his mobile device onto the doctor's mobile device to exchange his records using NFC NDEF format. The doctor can read and update the records and tap them back onto the patient's mobile device. Both patient and doctor register for the OPD session with the health admin, $Mobile_{ADMIN}$, to get secure keys. The patient's public and private keys Kp_{UBPAT}. Kp_{RJPAT} and doctor's public and private keys Kp_{UBDOC}, Kp_{RIDoc} get stored on the SE of their respective mobile devices for the OPD session, This Healthcard offers more storage space as compared to what a smartcard based tag can provide as in application III-B. It also ensures that only the permitted records of the patient are accessed by an authorized doctor, thus retaining security and privacy of the patient. NFC P2P mode can be utilized for information exchange, But very large health records exchanged over NFC can be slow due to the low data rate of NFC. Bluetooth can be used along with NFC for exchange of larger information.

D. e-Health Card based on NFC card emulation

In this application architecture, Healthcard is retained on a mobile device using card emulation and Java card applets installed on the SE. We propose usage of a SE in the form of an SWP enabled microSD card which can be issued to the patient by HealthSecure service. Java Card applet can be used to authenticate and authorize the reader to access and update the health records using NFC SWP protocol. Since the SE has limited space, it can only retain part of the health records. The remaining health records can be retained outside the SE region on the SD card in a secure manner. The Card on the $Mobile_{PAT}$ can be accessed externally by a PC/SC reader that is attached to $Mobile_{DOC}$. Since the SE has limited space, an extended card consisting of past records and other health information, like images and

reports, can be stored in encrypted format. Hence this Healthcard is different from a standard plastic smartcard used for Healthcard in the previous scenario. Since NFC has lower data rate, Bluetooth can be used to access the extended card. The Java card applet can be used to initiate Bluetooth pairing between mobile devices. This Healthcard is most secure and can also be used to retain larger information on the mobile device.

IV. IMPLEMENTATION

Applications have been developed for both Android devices using Android APIs, and administrative server, using PHP and MySQL, for secure, reliable and robust healthcare system. Mobile applications have been tested on latest android phones. We have used MIFARE Classic for reading and writing data using APIs in Android framework (Android 2.3.3 and above).The memory of this card is 1Kb. The Android framework provides android.nfc.tech package, which contains necessary classes and methods to enable interaction with tags. We have used a SWP Secure microSD card, which provides a microSD based Java Card 2.2.2 solution. The card supports Java Card applets on a hardware-backed SE. It also provides a contactless interface (ISO 14443) via SWP which can be used to interact with compliant PC/SC readers. We have tested it using an ASUS Zenfone 2 mobile device with Android 4.1.2. The card can be accessed from an authorized Android application. Since the card supports Global Platform 2.1.1, the installation can be done using custom Global Platform APDUs. Java card applets have been developed to store credentials for security framework and for card emulation mode. Implementation of Security framework and hybrid cloud service is in progress and will be tested and deployed in the field in our future work.

The healthcare data can be large in size as in a Health card with entire EHR in section. Also the health card could be accessed by various persons: patient, medical professional, emergency person and insurance. The patient should be able to securely manage the access control of the EHR. There is a requirement of confidentiality, integrity, mutual authentication, access control of EHR, privacy threats leading to identity thefts and insurance security breach. The security framework involves various entities. A cryptographic server is used to generate, verify and store security keys. An administrator is present to issue and authenticate Healthcards / tags and register patients/doctors. Mobile devices used by doctors are equipped with a Doctor App and a secure element. Healthcard used by patients is called Patient card which in this case is using a NFC P2P or card emulation mode. Since the health card could be accessed by various persons: patient, medical professional and emergency person. There could be a separate Doctor PIN for doctor and a super key for emergency team when patient is unconscious. The security flow consists of 1. Healthcard personalization. 2. Mutual Authentication between the patient and the medical doctor to assure the correct patient is appearing before an authorized doctor and there is no relay attack. 3. Access control for data viewable by the doctor. 4. Secure healthcard retrieval and updation. There is an initial phase of personalization in which the Patient Card and the Doctor get a unique identification ID (UID_{pat} and UID_{doc}) and a set of public and private keys (K_{pUBPAT}, K_{pRIPAT} and K_{pUBOOC}, K_{pRIDoc}) which are stored locally in the security server based on the respective card ID and/or secure element ID. An encrypted and signed data communications ensures confidentiality and integrity.

V. MEDICAL USE CASES OF NFC

There are a lot of use cases for NFC in medical devices and healthcare. The possible areas include monitoring and management of home based care. The application includes monitoring systems for a variety of chronic diseases, including but not limited to diabetes, hypertension, cardiac diseases (infarctions, heart failure, arrhythmias and other rhythm abnormalities), pulmonary diseases like asthma and COPD, and neurological abnormalities like seizures, chronic renal failure, etc. For example, a biometric device called "MiniME" developed by Ergonomidesign monitors various vital parameters like ECG, blood pressure, heart rate, pulse oximetry, body temperature, blood glucose, cholesterol, haemoglobin and prothrombin time, and transmits the data using NFC to the cloud.

Another company working on medical devices with NFC embedded in them is Impak Health. They are involved in home-based cardiac, pulmonary and sleep monitoring. They have incorporated NFC in devices such as "RhythmTrack" that tracks a person's ECG, and "SleepTrack," which tracks the sleep cycle and duration. Similarly, FITBIT – a fitness monitoring company – has incorporated NFC for transferring details like calories burned, number of steps taken and other details from a wristband to the user's smart phone which houses a user-friendly application. Gentag, a company specializing in mobile health, is using NFC to transfer data from devices ranging from diagnostic assays to skin patches. Nedap, a Netherlands-based security and identification specialist, has rolled out 50,000 NFC phones for nurses. They are used for recording home visits for the elderly. NFC is becoming widely accepted for medical devices in some markets specifically in the developed countries. Sony Corporation has developed an NFC Healthcare Library which enables communication between healthcare products embedded with the NFC Dynamic Tag (FeliCa Plug) and healthcare applications installed on smart phones. This library is available free of charge for a number of OS, including Windows, Linux and Android. Companies like Omron, Terumo and A&D are incorporating Sony's solution into their devices like BP monitors, pedometers, blood glucose meters, etc. Various other companies like Qolpac and Identive WPG have brought NFC into the mainstream with uses ranging from medication compliance to X-ray image sharing.

CONCLUSION

Near field communication can be extremely beneficial in the modern era of technology. It is interactive and secure which does not require any special software to run on. It is also more intuitive, making it a good candidate for use in the home-based monitoring and management domain, particularly among the elderly. We have proposed

applications based on NFC enabled Android mobile devices for improving healthcare process for secure medical object identification and patient Health card on an external tag or mobile device itself. The system has improved access to patients' medical history and improved medical checkups by automatically updating information and access to the entire patient's medical records. This will improve the health flow in crowded hospitals of developing countries as well as of developed nations. While it is still at the fringes and is waiting for its big break, NFC is being increasingly adopted by a number of organizations. Certain security-related concerns need to be allayed and the solutions developed around NFC have to be demonstrably secure. Taking all these factors into consideration, it is reasonable to conclude that NFC is a promising technology. Adoption of NFC in medical devices will help increase the security and ease of data transfer between medical devices and also provide personalized healthcare.

REFERENCES

[1] Divyashikha Sethia, Huzur Saran, **"NFC Based Secure Mobile Healthcare System",** IEEE Transactions on Communication Systems and Network(COMSNETS), Jan 2014.

[2] A Devendran, Dr. T Bhuvaneshwari, Arun Kumar Krishnan, **"Mobile Healthcare System using NFC technology",** International Journal of Computer Science Issues(IJCSI), May 2012.

[3] Ketan Bhadoriya, Rajshekar Humbe, Yogesh Kodgire, S.G. Phule, **"NFC Based Healthcare System",** International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6, 2015.

[4] Bankar Karthik, Joshi Bhargav, Mungal Mahajan, Subhash Rathod, **"NFC Based Android API Healthcare System",** Multidisciplinary Journal of Research in Engineering and Technology, March 2015.

[5] Prasad S. Halgoankar, Vijay M. Wadhai, **"NFC based Healthcare System for Mobile Prescription",** International Journal on emerging trends in Technology(IJETT), September 2015, Volume 2, Issue 2.

[6] Divyashikha Sethia, Daya Gupta, Tanuj Mittal, Ujjwal Arora, Huzur Saran, **"NFC Based Secure Mobile Healthcare System",** IEEE Transactions on Communication System and Network(COMSNETS), Jan 2014.

[7] Sanjana H.D, Gopika Priyadarshini, Yamini Gadidam, Shruthi B, Vikranth B.M, **"NFC Based Secure Mobile Healthcare System",** Journal of Emerging Technologies and Innovative Research(JETIR), April 2015, Volume 2, Issue 4.

[8] Danco Davcev, Goran Jakimovski, **"Ergonomics design of Healthcare NFC-based System",** International Conference on Applied Human Factors and Ergonomics (AHFE 2015).

Human Activity Recognition using Android Smartphone

Usharani J, Dr. Usha Sakthivel
PG Scholar, Professor and HOD, Department of Computer Science,RRCE, Bengaluru-74
usharanismg@gmail.com, sakthivelusha@gmail.com,

ABSTRACT- Activity recognition is one of the most important technology behind many applications such as medical research, human survey system and it is an active research topic in health care and smart homes. Smart phones are equipped with various built-in sensing platforms like accelerometer, gyroscope, GPS, compass sensor and barometer, we can design a system to capture the state of the user. Activity recognition system takes the raw sensor reading from mobile sensors as inputs and estimates a human motion activity using data mining and machine learning techniques. In this paper, we analyze the performance of two classification algorithms i.e. KNN and Clustered KNN in an online activity recognition system working on Android platforms and this system will supports on-line training and classification using the accelerometer data only. Usually first we use the KNN classification algorithm and next we utilize an improvement of Minimum Distance and K-Nearest Neighbor classification algorithms, i.e. Clustered KNN . For the purpose of activity recognition, clustered KNN will eliminates the computational complexities of KNN by creating clusters (creating smaller training sets for each actions and classification will be performed based on these reduced training sets). We can predict the performance of these classifiers from a series of observations on human activities like walking, running, lying down, sitting and standing in an online activity recognition system. In this paper, we are intended to analyze the performance of classifiers with limited training data and limited accessible memory on the phones compared to off-line.

Keywords - Accelerometer; Activity Recognition ; Clustered KNN; KNN; Smartphone.

I. INTRODUCTION

Nowadays smartphones became more and more popular in human daily life. Most of the people used it for searching news, watching videos, playing games and accessing social network but there were many useful studies on smartphones. Activity recognition is one of the most important technologies behind many applications on smartphone such as health monitoring, fall detection, context-aware mobile applications, human survey system and home automation etc., Smartphone-based activity recognition system is an active area of research because they can lead to new types of mobile applications.

Understanding human activities creating a demand in health-care domain, especially in rehabilitation assistance, physiotherapist assistance, and elder care support services and cognitive impairment. Sensors will record and monitor the patient's activities and report automatically when any abnormality is detected, so, huge amount of resources can be saved. Other applications like human survey system and location indicator are all getting benefits from this study.

Training process is always necessary when a new activity is added in to the system. The same algorithm parameters are needed to be trained and fine-tuned when the algorithm runs on different devices with various built-in sensors. However, labeling a training data (time-series data) is a time consuming procedure and it is not always possible to label all the training data by the users. As a result, we present an active learning technique to accelerate the training process. Given a KNN classifier, an active learning technique intuitively queries the unlabeled training samples and learns the parameters from the correct labels answered by the human. In this way, users will label only the samples that the algorithm demanded to do and the total amount of required training samples is reduced.

HAR system takes the raw sensor readings from mobile sensors as inputs and predicts human motion activity, this can be done by leveraging smartphone with various sensors, including accelerometers, compass sensor, GPS, light sensors, gyroscope, barometer etc., Due to its unassertive, none/low installation cost and easy-to-use, smart phones are becoming the main platform for human activity recognition**.** In this paper, we focus on robust human activity recognition using 3-dimentional accelerometer and gyroscope on smart phones.

In this paper, we are also interested in analyzing the performance of classifiers with limited training data considering the limited memory available on the phones. In this system, we can collect the training data in a few minutes and it can be directly used for classification steps, which reduce the burden on the users. Being one of the first Android applications used for human activity recognition is another important motivation for this study. In the literature, it has been reported that minimum distance classifier does not work well when used alone. KNN results are always better than minimum distance classifier in terms of accuracy. However, KNN requires high computational burden so, it is not an online classifier and due to limited resources on smart phone, it does not appear as a preferable method.

The rest of the paper is organized as follows. We describe Related work in Section II and Section III will describes the Android Smartphone sensors used in activity recognition system. Section IV describes Core Techniques and the human activity recognition on Android Smartphone is described in Section V. Result and Discussion is described in Section VI. Finally, we conclude the paper in section VII.

II. RELATED WORK

Human activity recognition on smartphone is an active research area. Most of related works focus on analyzing the performance of classification algorithms such as: Decision Trees, Naïve Bayes, Nearest Neighbor algorithms, Support Vector Machines, Hidden Markov Chain, Multi-Layer Perceptron and Random Forrest. There are not much work investigated on features selection of dataset. Jun Yang and et al. extracted orientation-independent features from three feature sets, including horizontal, vertical and magnitude features. Each feature set consists of mean, standard deviation, zero cross rate, 75 percentile, interquartile, spectrum centroid, entropy. The authors used Attributed Selection filters to give 7 feature subsets and evaluate recognition accuracy on these subsets. As a result, the accuracy of classifiers on each subsets are lower than with all features, i.e. Decision Tree equals to 90.4% (all features: 90.6%), Naïve Bayes equals to 68.3% (all features: 68.7%). Sian Lun Lau and et al. used common four features mean, standard deviation, energy of the Fast Fourier Transform and correlation. They combined features to 3 groups: group G1 includes average and standard deviation of values of each axis and all three axes, group G2 includes the average and the standard deviation of FFT coefficients of each axis and all three axes, group G3 includes all four features of each axis and all three axes. However, they used simple features and combined features into groups manually. Ville Kononen and et al. used two feature selection methods, including Sequential Forward Selection and Selection to select features from accelerometer and heart rate signals and evaluate complex classification compared with simple classification. However, they used the feature selection method to select features and compared accuracy of classifier on that features and recognition accuracy of classifier range from about 60% to 90%.

Different from other work, in this paper we do not only remove the irrelevant and redundant features but also remove redundant instances. As a result, the system will achieve better recognition whereas the training dataset is reduced significantly.

III. ANDROID SMARTPHONE SENSORS

Android devices have built-in sensors that measures motion (accelerometers, gravity sensors, and gyroscope), orientation (magnetometers.), and various environmental conditions (barometers, photometers, and thermometers). These sensors are capable of providing raw data with high accuracy, and are useful if we wanted to monitor three-dimensional device movement or positioning, or we wanted to monitor changes in the ambient environment near a device.

The Android sensor framework allows us to access many types of sensors, such as hardware-based and software-based sensors. Hardware-based sensors are physical components built into a smartphone and they derive their raw data by directly measuring specific environmental parameters, such as acceleration, geomagnetic field strength, or angular change. Whereas, software-based sensors (linear acceleration sensor and the gravity sensor) are not physical devices, although they imitate hardware-based sensors. Software-based sensors derive their raw data from one or more hardware-based sensors and are sometimes called virtual sensors or synthetic sensors.

The android sensor framework uses a standard 3-axis coordinate system to express data values. For most sensors, the coordinate system is defined relative to the device's screen when the device is held in its default orientation (see fig 1). When a device is held in its default orientation, the X axis is horizontal and points to the right, the Y axis is vertical and points up, and the Z axis points toward the outside of the screen face. In this system, coordinates behind the screen have negative Z values.

1 Accelerometer

Accelerometer is a latest technology which has upgraded the user experience in smartphones and it measures the acceleration force applied to a device on all three physical axis (x, y, and z), including the force of gravity. It changes the orientation and adjusts the screen to proper viewing, when user changes the orientation from landscape/horizontal to portrait/vertical and vice-versa. The smartphone physical position can be determined by 3-way axis device. Fig 1 shows the Accelerometer axes on smartphone. The raw data from the accelerometer is represented in a set of vectors: Acc i=<x i, y i, z i> , where i=(1,2,3,….). A time stamp can also be returned with these 3-axis readings.

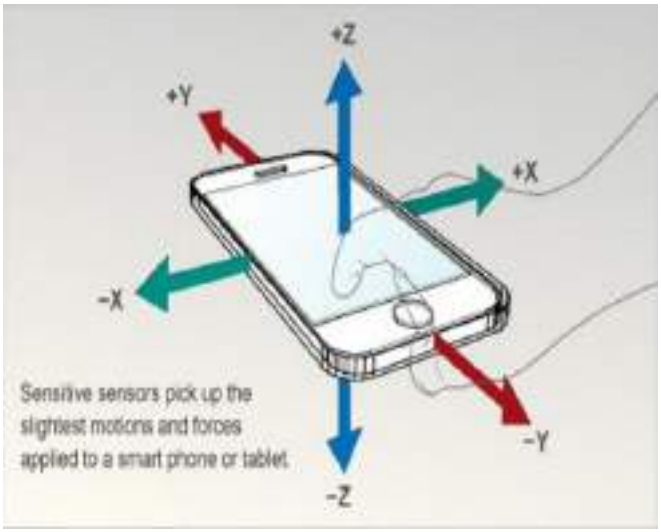


Fig 1: Accelerometer axes on smartphone.

2 Compass sensor

The digital Compass is a traditional tool to detect the direction with respect to the north-south pole of the earth's magnetic field. Compass functionality in smartphones is usually based on more sophisticated sensor called a magnetometer; it is used to measure the strength and direction of magnetic fields. Fig 2 shows the compass reading display screen on a Smartphone. By analyzing Earth's magnetic field, the sensor allows a phone to determine its orientation with high accuracy. The raw data reading from a compass sensor is the float number between 0º and 360º. It begins from 0º as the absolute north and the

actual reading indicates the angle between present smart phone direction and the absolute north in clockwise. The Pro version of Smart Compass adds a speedometer and the option to send GPS coordinates via SMS or email. Compass reading can be used to detect the direction change in the human motion such as walking.

Fig 2 : Compass Sensor on Smartphone's.

3 Gyroscope

The gyroscope is a device, which adds an additional dimension to the information supplied by the accelerometer by tracking rotation or twist and it is primarily used for navigation and measurement of the angular rotational velocity. And it also uses earth's gravity to help determine orientation. Gyroscope measures the phone's rotation rate by detecting the roll, pitch, and raw motions of the smart phones along the x, y, and z axis, respectively. The axes directions are shown in Fig. 3. The raw data from a gyroscope is the rate of the rotation in rad/s (radian per second) around each of the three physical axes: Rotation i =<xi ; y i ; z i >; I =(1,2,3,.....) . In activity recognition search, gyroscope is used to assist the mobile orientation detection.

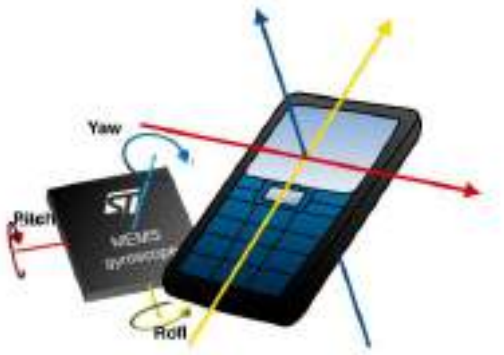


Fig 3: Gyroscope Three Axes on Smartphone's.

4 Barometer

Barometer is a device equipped on most of the advanced smart phones. It measures the atmospheric pressure of the environment where the sensor is placed in. So, barometer reading can be used to indicate the user's position change in localization related activity recognition. Barometers senses air pressure and are used in smartphones to determine relative elevation - measuring stairs climbed and so on. It should be able to take air pressure measurements and then help forecasters analyze where troughs, high-pressure zones, and frontal boundaries are.

IV. CORE TECHNIQUES

1 KNN

K-Nearest Neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest Neighbor category and it is one of the most popular algorithms for activity recognition. The purpose of KNN algorithm is to stores all available objects and classifies a new object based on attributes and training samples. The KNN classifiers do not use any model to fit and only based on memory. KNN algorithm used neighborhood classification as the prediction value of the new query instance. Fig 4 shows KNN Algorithm.

In activity recognition, the k-nearest neighbour algorithm (K-NN) is a non-parametric (or distribution free) method for classifying objects based on closest training samples in the feature space. KNN is a type of memory-based learning, where the function is only approximated locally and all computation is delayed until classification. The KNN algorithm is amongst the simplest of all machine learning algorithms: here, an object is classified by a majority vote of its neighbors, with the object being assigned to the class amongst its k nearest neighbors (k is small positive integer). If k=1, then the object is merely assigned to its nearest neighbor class. The training examples are vectors in a multidimensional feature space, each with a labeled class. The training (data pre-processing) phase of the algorithm consists storing the feature vectors and class labels of the training samples. In the classification phase of activity recognition system, k is a user-defined constant, and an unlabeled vector is classified by assigning most frequent label among the k training samples.

Usually Euclidean distance is used as the distance metric for continuous variables;

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Hamming distance is used as distance metric for discrete variables, such as text classification.

$$D_H = \sum_{i=1}^{k} |Xi - Yi|$$

$$X = Y \rightarrow D=0$$

$$X \neq Y \rightarrow D=1$$

The classification accuracy of *K*NN can be improved significantly if the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighboror or Neighbourhood components analysis.

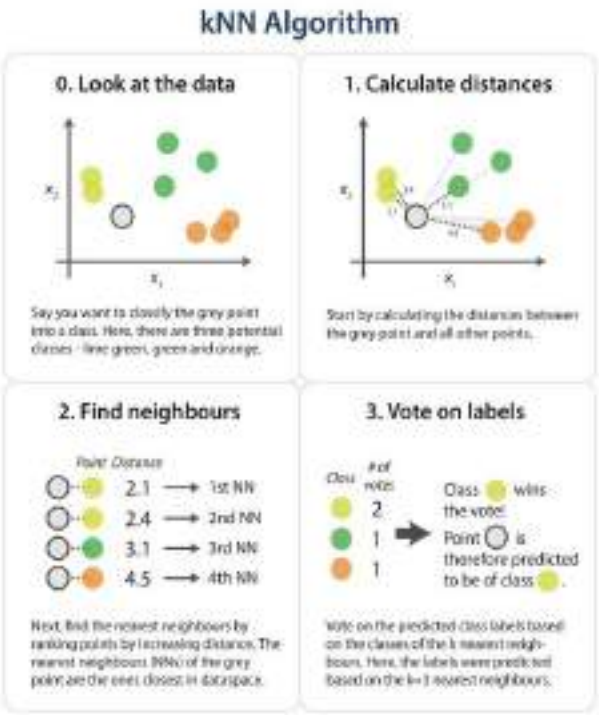


Fig 4 : KNN Algorithm.

2 Clustered KNN

K-Nearest Neighbor classifier may have a problem when training samples are uneven. The problem is that KNN classifier may decrease the accuracy of classification because of the uneven density of training data. To solve this problem, a new clustering-based KNN method is presented in this paper. In the first step the training data is pre-processed and four features, which are average, minimum, maximum, and standard deviation, are extracted by using clustering and in the second step classification takes place with a new KNN algorithm; which adopts a dynamic adjustment in each iteration for the neighborhood number K.

Clustering is a process of partitioning data into clusters of similar objects. It is an unsupervised learning process of hidden data.

i) Data Pre-processing in Clustered KNN

The main objective of the pre-processing step is to define activity sets from the training data based on the mentioned features. Instead of comparing all the data in the training set, we compare the test data only with the compact training data set that we selected from the original training set. During the pre-processing step, for each feature and for each activity, compact training sets are created. For each feature, except the standard deviation, K data points are selected from the training data. For instance, for the minimum feature set, K "minimum" data points are selected from the training data. Likewise, we create a "maximum" set by selecting the K maximum data points. The average value of the training data is calculated and the nearest K data points are included in the "average" set. For the "standard deviation" set, standard deviation value of training data for each activity is calculated. However, at the same time accuracy of the results are expected to decrease with smaller value of K, so that there is an important trade-off between accuracy and execution time considering the value for K.

ii) Classification in Clustered KNN

In the classification step, we collect test data, in other words we segment the data during a window with a predefined size. After the window is filled, classification phase will starts, and average, minimum, maximum, standard deviation values of the data in the window is calculated and these values are compared one by one with the values in the compact training sets which were created during the pre-processing step. *K* nearest sample to test data is selected from training sets and voting is done by looking at the final list of activities. We label the data in the related window as the activity for which we have maximum amount of data in the final *K* set. The one which is closer to the standard deviation of that particular window is selected as the recognized activity by the standard deviation feature. At the end, we have four labels from voting results of each feature. We label the window as the activity for which we have the highest vote and finalize the classification.

V. HUMAN ACTIVITY RECOGNITION ON ANDROID PLATFORM

This section describes the human activity recognition system based on smartphone, The clustered KNN classifier are implemented on Android phones to detect five main activities; which are walking, running, lying down, standing and sitting. For this purpose, the process is divided into two phases: Data pre-processing and Classification as shown in Fig 5.

In order to monitor the performance of the classifier, the ground truth data is logged, i.e. which activity is literally performed by the user. For this purpose, the application gives voice commands repeatedly to perform an activity. The order of activity is predefined in the system whereas activity duration "order interval" is given directly as the user input to the system in the unit of seconds. During our experiments each activity is performed for 60 sec/cycle. Finally, using these ground truth values, i.e., activity tags, activity recognition performance and other performance metrics of the classifiers are calculated.

Phase I : Data Pre-processing

1) Data collection

Data is collected through the application we created called activity logger and this component is responsible for collecting the training data for each activity separately from sensors, i.e. accelerometer data. In this application, user will selects the activity to be performed, keep the phone into the packet and starts to perform the related activity. For each activity, this application is responsible for creating different training data files in which raw data from the 3-axes of the accelerometer is being logged.

2) Data filtering

The collected data may contain noise and the data collection is processed the data to eliminate the noise by applying noise filters and low-pass filter technique. Time-domain and frequency-domain features have been extensively used to filter relevant information within acceleration and rotation signals. In this paper, we used four features : MIN, MAX, MEAN and Standard deviation. By these statistical operations features were calculated.

3) Data segmentation

Data segmentation is a crucial stage in the activity recognition process; normally sliding window approach is used for segmentation but no clear consent exists on which window size should be preferably employed. Intuitively, decreasing the window size allows faster activity detection, as well as reduced resources and energy needs. On the contrary, large windows are usually considered for the recognition of complex activities. The filtered sensor data is divided into small segments for feature extraction using windowing approaches.

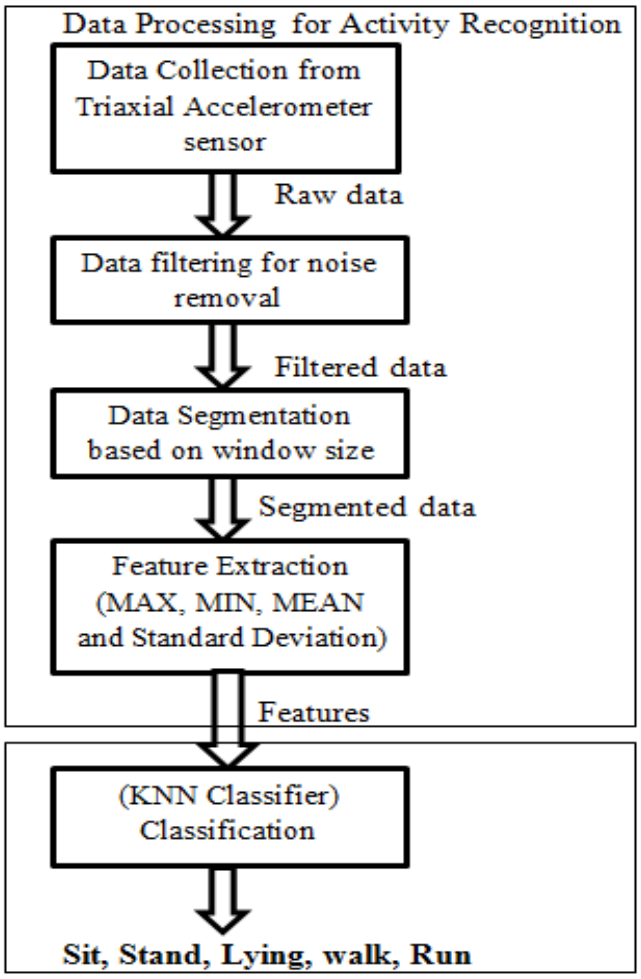


Fig 5 : Human Activity Recognition System

4) Feature Extraction

As in any other data mining tasks, extracting the 'right' features is critical to the final recognition performance. For activity recognition, we can extract features in both time and frequency domains.

Feature extraction is an most important pre-processing step to activity recognition. It is often decomposed into feature construction and feature selection. When the input data to an algorithm is too large to be processed and it is suspected to be useless then the input data will be transformed into a reduced representation set of features. Transforming the input data into set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the most relevant information from the input data in order to perform the specified task using this reduced representation set of features instead of the full size input. Feature extraction is performed on raw data prior to applying *K*NN algorithm on the transformed data in feature space.

a) Dimensionality Reduction

As shown in Figure 4.1, the Dimensionality Reduction is responsible for reducing the computational complexity thus reducing the response time for recognition process, whereas ensuring accuracy of recognition. It contains two components: the Feature Selection Component and the Instance Selection Component. The Feature Selection is responsible for identifying and removing unneeded, irrelevant and redundant attributes from the dataset. The Instance Selection is responsible for removing unneeded, irrelevant and redundant instances from dataset.

b) Feature Selection

The Feature selection will reduce the processing cost by remove irrelevant and redundant features, whereas ensuring the accuracy of recognition. The redundant features are those which do not contribute information to the recognition process. In other words, the features which provide no/irrelevant to the selected current features set should be eliminated. As a result, it will improve model interpretability, shorter training time, and enhance generalization by reducing over fitting.

Phase II: Classification

When the window is filled, classification phase will starts, and average, minimum, maximum, standard deviation values of the data in the window is calculated and these values are compared one by one with the values in the compact training sets which were created during the pre-processing step. *K* nearest sample to test data is selected from training sets and voting is done by looking at the final list of activities. We label the data in the related window as the activity for which we have maximum amount of data in the final *K* set. The one which is closer to the standard deviation of that particular window is selected as the recognized activity by the standard deviation feature. At the end, we will get four labels from voting results of each feature. We label the window as the activity for which we have the highest vote and finalize the classification.

VI. RESULT AND DISCUSSION

To evaluate the performance of clustered KNN classification for each activity. The confusion matrix for clustered KNN is presented in Table 1. Confusion matrix is a visualization tool typically used in supervised learning techniques. One advantage of confusion matrix is that it is easy to see if the system is confusing two classes. Each column of the confusion matrix represents the instances in a predicted class, while each row represents the instances in an actual class. A confusion matrix contains information about known class labels and predicted class labels. Compared to the performance of activities like running, lying down, standing and sitting, the KNN classifier presents slightly worse performance for walking, where this activity is sometimes classified as running or standing. However, the overall performance for clustered KNN classification is around 92% accuracy considering all activities.

We are also evaluated the impact of *K* value on the performance of clustered KNN classification for each activity. As we expected, increasing the *K* value will affects accuracy rates positively. Whenever we consider the overall effect of all system parameters, we observe best results in the case where K is selected as 50, window size is selected as 1 second and sampling interval is selected as 50 msec.

Table 1: Confusion Matrix

Window size(sec)		0.5			1			2		
Sampling Interval(msec)		10	50	100	10	50	100	10	50	100
K	10	87.9	87.8	87.7	88.4	90.3	89.5	88.6	87.8	89.3
	50	91.1	90.0	91.4	91.9	92.1	90.8	90.9	89.4	91.0

VII. CONCLUSION

In this paper, we proposed an activity recognition system working on Android platforms by developing an application called application logger that supports on-line training and classification while using only the accelerometer data for classification. The performance of on-line classification of KNN classifier is evaluated first then a clustered KNN method is used. The clustered KNN classification exhibit a much better performance than the KNN classifier in terms of accuracy on android platforms with limited resources. We also evaluated the performance of clustered KNN in terms of execution times. As we expected, classification execution times are considerably reduced as K parameter is decreased. Additionally, classification times are highly dependent on the device model and capabilities as well.

Future work will evaluate whether additional features are necessary to improve classifier performance, without adding computational complexity to the algorithm and we will investigate the performance in more complex activities recognition such as bicycling, fall detection.

References

[1] Xing Su, Hanghang Tong, and Ping Ji , "Activity Recognition with Smartphone Sensors" ,tsinghua science and technology ISSN ll1007- 0214ll02/11llpp235-249 Volume 19, Number 3, June 2014.

[2] Nicholas D.Lane , Emiliano Miluzzo, Hong Lu, Daniel Peebles,"A Survey of Mobile Phone Sensing",IEEE Communications Magazine September 2010

[3] Sahak Kaghyan, Hakob Sarukhanyan, "Activity recognition using knearest neighbor algorithm on smartphone with tri-axial accelerometer ",International Journal "Information Models and Analyses" Vol.1 / 2012J.

[4] Mustafa Kose, Ozlem Durmaz Incel,Cem Ersoy "Online Human Activity Recognition on Smart Phones"2nd International Workshop on Mobile Sensing, April 16, 2012.

[5] Pekka Sirrtola and Juha Röning, "Recognizing Human Activities Userindependently on Smartphones Based on Accelerometer Data",International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 1, No 5.

[6] Jennifer R.Kwapisz Gary M. Weiss, Samuel A. Moore,"Activity Recognition using Cell Phone Accelerometers",sensor KDD '10,july 25,2010 ,Washington,DC,USA.copyright 2010ACM.

[7] Ling bao and stephen S. Intille,"Activity Recognition from User-Annotated Acceleration Data",. Ferscha and F.Mattern (Eds) PERVASIVE 2004, LNCS 3001, pp. 1–17, 2004. Springer-Verlag Berlin Heidelberg 2004.

[8] Won-Jae Yi, Weidi Jia, and Jafar Saniie ,"Mobile Sensor Data Collector using Android smartphone", Department of Electrical and computer Engineering illinos institute of technology 3301 S.Dearborn St.103SH,Chicago IL,USA.

[9] I. Anderson, J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, B. Brown, and H. Muller. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. Mobile Networks and Applications, 12(2-3):185–199, 2007.

[10] Android. Sensors Overview. http:// developer.android.com/ guide/ topics/ sensors/ sensors_overview.html, 2014. [Online; accessed 01-March-2014].

[11] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In Ambient Assisted Living and Home Care, pages 216–223. Springer, 2012.

[12] Apple. Nike + iPod Application, 2014. [Online; accessed 01-March-2014].

[13] Apple. UIAcceleration Class Reference. https://developer.apple.com/library/ios/documentation/uikit/reference/UIAcceleration_Class/Reference/UIAcceleration.html, 2014. [Online; accessed 17-March-2014].

[14] W.-Y. Deng, Q.-H. Zheng, and Z.-M. Wang. Cross-person activity recognition using reduced kernel extreme learning machine. Neural Network

\

A Reward-based MAC layer scheduling For Co-operative Multi-Hop Cognitive Radio Networks

Haseeba Yaseen, Snigdha Sen
Asst.Prof, Department of Computer Science, Global Academy of Technology, Bangalore-98
haseebayaseen@gmail.com, snigdha.sen@gat.ac.in

ABSTRACT- Cognitive Radio Networks (CRNs) have recently emerged as a technology for secondary users (SUs) to opportunistically utilize the spectrum assigned to primary users (PUs). The purpose of this paper is to design a mac-layer optimal scheduling algorithm for cooperative multi-hop Cognitive Radio Networks (CRNs), where SUs assist PUs multi-hop transmissions and in return gain an immediate time-share of the channel proportional to their assistance.A time slotted multi hop co-operative CRN is where SUs relay PU data in return for the right to use the wireless spectrum is considered. The multi hop co-operative CRN is divided into two sub networks; a PU relay sub network and an SU sub network. The throughput optimal scheduling algorithm is designed with two mechanisms; Immediate- Reward Mechanism and Long-term Reward Mechanism. The algorithm is composed of two parts, namely, a congestion controller and a hop/link scheduler. The properties of proposed algorithm are illustrated through simulation studies.

Keywords - Cognitive Radio, Congestion Control, Finite Buffer, Multi-Hop Cognitive Radio Networks, Optimal Scheduling..

I. INTRODUCTION

Cognitive Radio Networks (CRNs) [3] have recently emerged as a new technology for unlicensed users to utilize the under-used spectrum opportunities. In a typical CRN, licensed users are referred to as primary users (PUs) and secondary users (SUs) denote the users dynamically utilizing spectrum opportunities. The concept of CRN is simple, but the design of CRNs imposes challenges that are not present in conventional wireless networks [3]. The traditional view on CRNs emphasizes point to- point connections for both PU and SU subsystems, and multi-hop CRNs have only been considered in recent past.

In this paper, a throughput-optimal mac-layer scheduling algorithm for a multi-hop cooperative CRN under a property-rights model [8] is proposed,where SUs relay data between PU pairs to gain access to the licensed spectrum. An illustrative example is shown in Figure 1, where the cooperative CRN is composed of an SU sub network and a PU sub network. The SU sub network consists of SUs communicating with a secondary base station over a single hop as assumed for IEEE 802.22. In the PU sub network, we consider a case where the channel condition is not desirable for the direct transmission between the PU and the primary base station due to physical separation. Thus, the PU is willing to "lease" a portion of the spectrum access to SUs in return for some form of service. Specifically, PU data is relayed by SUs from the source PU to PU base station, and SUs in return gain a time-share of the channel proportional to their assistance to the PU. The model illustrated in Figure 1 can be considered as a generalization of the overlay CRNs with two-hop relay [8]-[7].

In the proposed algorithm: the SUs are guaranteed a throughput proportional to the PU data they relay. An optimal opportunistic scheduling scheme has been proposed in [7] to guarantee each user a proportional share of the network resource for a non-cognitive setting, which is extended to scenario of two-hop relay CRNs [8].

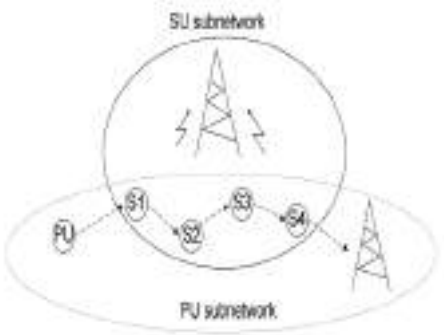


Fig.1: Co-operative CRN Model

II. NETWORK MODEL

2.1 The Cognitive Network

Consider a cognitive radio network of M primary users and N secondary users, all wishing to communicate with a common destination as shown in figure 2. This destination can be viewed as a base station in a single-cell of a cellular network or as an access point in a Wi-Fi network. A time-slotted system where the time slot is the resource to be shared among different nodes is considered and a non-interference model where only one node, either primary or secondary, is transmitting at any given time1 is adopted.

Fig.2: Network Model

Random channel gains between each node and other nodes in the network are assumed to be independent and identically distributed across time according to a general distribution and independent across users with values taken from a finite set.

2.2 Overall Network Elements and Constraints

In this paper, a time-slotted multi-hop cooperative CRN, as illustrated in Figure 1 is considered, where SUs relay PU data in return for the right to use the wireless spectrum. The multi-hop cooperative CRN in question can be divided into two sub networks: a "PU relay sub network" and an "SU sub network". The PU relay sub network is composed of one primary source node (*sP*), a corresponding primary destination node (*dP*) which is represented as a primary base station in Figure 1, and a set of SUs *S* that relay the PU traffic between *sP* and *dP* over possibly multiple hops, where | *S* |= *N*. This model can be considered as a generalization of the overlay CRNs with two-hop relay [5]-[7] and assumed that *sP* and *dP* cannot communicate directly. Thus, PU data is relayed solely by SUs.

The PU relay sub network is represented as (*N*, *L*) where *N* = {*sP* , *dP*} ∪ *S* denotes the node set of the PU relay sub network and *L* denotes the link set for PU data relay, i.e., *L* = {(*m*, *n*) : *m*, *n* ∈ *N*, and there exists a link between nodes *m* and *n*}. The SU sub network is composed of a set of SUs *S* and secondary base station dS as their one hop destination. Then, the SU sub network can be represented by (*S* ∪ {*dS*},*L*_), where *L*_ = {(*l*, *dS*): *l* ∈ *S*} is the set of uplinks in the SU sub network. Let *V* = *L* ∪ *L*_. The CRN interference model is represented by an interference graph *G* = (*V*, *E*), a pair of links in *V* is in *E* if the links interfere with each other when scheduled simultaneously. Let μmn be the scheduled link rate for PU data over link (*m*, *n*) ∈ *L*, and the scheduled SU link rate denoted as *sl* over link (*l*, *dS*) ∈ *L*_. For analytical simplicity, a scheduled link rate takes a value from {0, 1} in units of packets per time slot.

A link schedule represented by a vector $((\mu mn)(m,n)\in L, (sl)l\in S) \in \{0, 1\}^{|L|+N}$ is said to be *feasible* iff any pair of scheduled links does not belong to the interference edge set *E*. With a time slot system, a feasible link scheduler chooses a feasible link schedule $((\mu mn(t))(m,n)\in L, (sl(t))l\in S) \in I$ for each time slot *t*, where *I* is the set of all feasible link schedules.

III. PROPOSED ALGORITHM FOR THE CRN

In this section, the throughput-optimal scheduling algorithm with the immediate reward mechanism [1] is described; the algorithm is composed of two parts, namely, a congestion controller and a hop/link scheduler. The formalized algorithm description is provided in Figure 3. The congestion controller generates and admits PU packets into the PU relay sub network, and a corresponding fraction of SU packets are admitted to their sources according to the immediate reward mechanism. The hop/link scheduler regulates the link transmission rates of the cooperative CRN.

Step 1. Congestion Controller:

$k^* = \arg\min_k \left(\rho_k \sum_{l\in P_k} Q_l(t) + U_0^k(t)\right)$;

for $k = 1, \ldots, K$

if $k = k^*$ do:

if $\rho_{k^*} \sum_{l\in P_{k^*}} Q_l(t) + U_0^{k^*}(t) \le V_2$ do:

$\mu^k_{-1,0}(t) = \mu_M$;

for $m \in \{1, \ldots, H_k\}$

do: admit $\rho_k \mu_M$ packets to SU v_k^m;

end for

else do: $\mu^k_{-1,0}(t) = 0$;

end if

else do: $\mu^k_{-1,0}(t) = 0$;

end if

end for

Step 2. Hop/Link Scheduler:

Find $\{(\mu_{mn}(t))_{(m,n)\in\mathcal{L}}, (s_l(t))_{l\in S}\} \in \mathcal{I}$ that maximizes:

$\sum_{k=1}^{K} \sum_{m=0}^{H_k} \mu^k_{m,m+1}(t)(U^k_m(t) - U^k_{m+1}(t)) + \sum_{l\in S} Q_l(t) s_l(t)$;

Update $(U^k_m(t+1))$ and $(Q_l(t+1))$ according to (3)(4).

Fig.3: Optimal algorithm with immediate reward mechanism in time slot.

3.1 Congestion Controller:

$$\min \sum_{k=1}^{K} \mu^k_{-1,0}(t)\left(\rho_k \sum_{l\in P_k} Q_l(t) + U_0^k(t) - V_2\right)$$

$$\text{s.t.} \sum_{k=1}^{K} \mu^k_{-1,0}(t) \le \mu_M.$$

The congestion controller (7) is a threshold based optimization problem, with the control parameter $V2$ as the threshold. The congestion controller (7) is developed to deterministically upper -bound the PU buffer size. Specifically, we will show later that $V2$ determines the finite PU buffer size and tradeoffs between throughput optimality and delay performance. For time slot t, we define

$k^* \triangleq \arg\min_k \{\rho_k \sum_{l\in P_k} Q_l(t) + U_0^k(t)\}$. Then, to solve (7), we set

$$\mu_{-1,0}^{k^*}(t) = \begin{cases} \mu_M, & \text{if } \rho_{k^*} \sum_{l\in P_{k^*}} Q_l(t) + U_0^{k^*}(t) \le V_2, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

For $k \neq k^*$, we set $\mu_{-1,0}^k(t) = 0$. With a centralized control, $\rho_k \mu_{-1,0}^k(t)$ packets are admitted to SUs c_k^m, $m = 1, \cdots, H_k$.

3.2 Hop/Link Scheduler:

In the hop/link scheduler (9), each SU link rate is weighted by the SU queue backlog and each PU hop rate is weighted by a "hop back-pressure", i.e., the difference between the PU queue backlogs across a hop where the optimization is taken over all feasible ((μk m,m+1(t))m,k,(sl(t))l∈S.

$$\max\Big\{\sum_{k=1}^{K}\sum_{m=0}^{H_k} \mu_{m,m+1}^k(t)\big(U_m^k(t) - U_{m+1}^k(t)\big) + \sum_{l\in\mathcal{S}} Q_l(t)s_l(t)\Big\}, \quad (9)$$

$$\text{s.t. } \{(\mu_{mn}(t))_{(m,n)\in\mathcal{L}}, (s_l(t))_{l\in\mathcal{S}}\} \in \mathcal{I},$$

The structure of the hop/link scheduler favors hops/links with higher weights for resource allocation, where we note that a higher weight implies a higher congestion level for a hop/link. When the hop back-pressure $Ukm(t) - Ukm+1(t) \le 0$, $m \in \{0, \cdots, Hk\}$, we set $\mu k\, m,m+1(t) = 0$, without loss of optimality.

IV. FURTHER DISCUSSION

With the immediate reward mechanism, the optimal backpressure- based algorithm requires simultaneous admission of both PU and SU packets. This requirement of simultaneous admission can be relaxed for an optimal algorithm with a long-term reward mechanism.

4.1 Proposed Algorithm with A Long-Term Reward Mechanism

In the original algorithm proposed in Section III with the immediate reward mechanism, SUs are assigned a channel share *proportional to* the relayed PU data, i.e., there may exist additional unutilized channel opportunities left by the PU. In addition, the congestion controller (7) is centralized to simultaneously admit both PU and SU packets. In this section, we extend our analysis to a CRN model with a more general ***long-term reward mechanism*** [1].The formalized algorithm description is provided in Figure 4.

4.2 PU Congestion Controller:

Redefining $k* \triangleq \arg\min k(\rho k\_\ l\in Pk\ Dl(t)+Uk\ 0(t))$, we admit the PU packets on the $k*$-th route as follows

$$\mu_{-1,0}^{k^*}(t) = \begin{cases} \mu_M, & \text{if } \rho_{k^*} \sum_{l\in P_{k^*}} D_l(t) + U_0^{k^*}(t) \le V_2, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

where $V2$ is the same control parameter as in Section III. For route $k \_= k*$, we set $\mu k-1,0(t) = 0$. Compared to the original congestion controller (7), we utilize the virtual queue $Dl(t)$ instead of the actual SU queue backlog $Ql(t)$.

Step 1. PU Congestion Controller:
$k^* = \arg\min_k \left(\rho_k \sum_{l\in P_k} D_l(t) + U_0^k(t)\right)$;
for $k = 1, \ldots, K$
  if $k = k^*$ **do:**
    if $\rho_{k^*} \sum_{l\in P_{k^*}} D_l(t) + U_0^{k^*}(t) \le V_2$
      do: $\mu_{-1,0}^k(t) = \mu_M$;
    else do: $\mu_{-1,0}^k(t) = 0$;
    end if
  else do: $\mu_{-1,0}^k(t) = 0$;
  end if
end for
Step 2. SU Congestion Controller:
for SU $l \in \mathcal{S}$ **do:**
  if $Q_l(t) \le D_l(t)$
    do: $A_l(t) = A_M$;
  else do: $A_l(t) = 0$;
  end if
end for
Step 3. Hop/Link Scheduler:
Find $\{(\mu_{mn}(t))_{(m,n)\in\mathcal{L}}, (s_l(t))_{l\in\mathcal{S}}\} \in \mathcal{I}$ that maximizes:
$\sum_{k=1}^{K}\sum_{m=0}^{H_k} \mu_{m,m+1}^k(t)(U_m^k(t) - U_{m+1}^k(t)) + \sum_{l\in\mathcal{S}} Q_l(t)s_l(t)$;
Update $(U_m^k(t+1))$ and $(Q_l(t+1))$ according to (3)(14).

Fig. 4 :Optimal algorithm with Long-term reward mechanism in time slot t.

4.3 SU Congestion Controller:

For each SU $l \in S$, The threshold-based SU congestion controller has a time varying threshold, i.e, the virtual queue $Dl(t)$.

$$A_l(t) = \begin{cases} A_M, & \text{if } Q_l(t) \le D_l(t), \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

4.4 Hop/Link Scheduler:

The hop/link scheduler remains the same as (9).

V. MEDICAL USE CASES OF NFC

In this section, we present a simulation-based performance evaluation for the algorithm proposed in Section III. Simulation results are obtained using the topology shown in Figure 1, which consists of a PU source (sP) and a PU destination (dP). In Figure 5, by fixing $V2 = 10$, we illustrate the throughput and congestion level performance of the algorithm against the route-specific reward parameters $\rho 1 = \rho 2 = \rho$, where we recall that the number of admitted secondary packets for each SU is ρ times the admitted PU packets and note that SU throughput is the sum for all SUs.

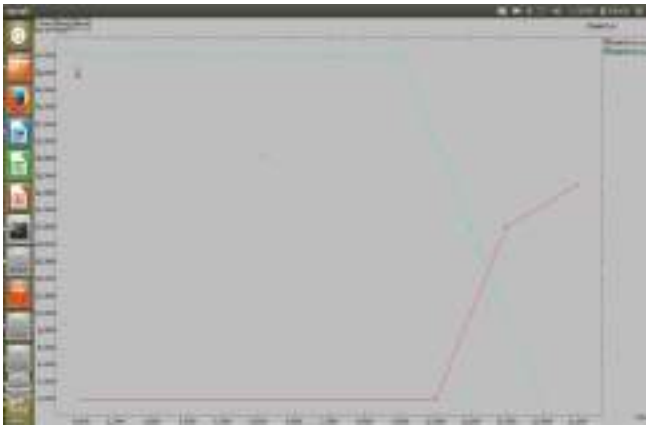


Fig. 5: Comparison of PU and SU Congestion.

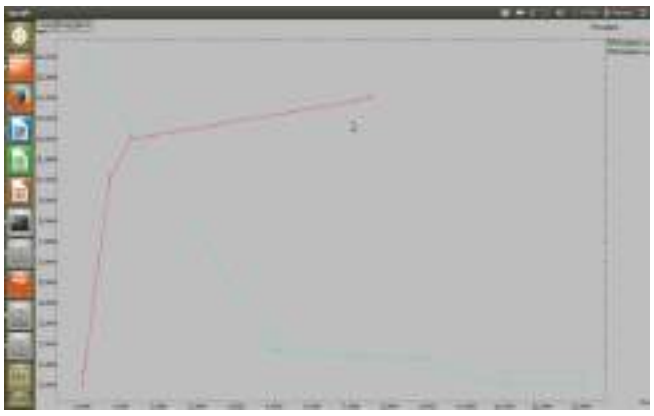


Fig. 6: Comparison of PU and SU Throughput.

According to the topology and the immediate reward mechanism, we must have the following relation between PU and SU throughput:

SU throughput= $2\rho \times$ (PU throughput), (18)

Noting that there are 2 SUs along each pre-determined route.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, a cross-layer scheduling algorithm for multi-hop cooperative cognitive radio networks is introduced. The algorithm can achieve a PU throughput arbitrarily close to the optimum, with a trade-off in the deterministically upper bounded PU buffer sizes. The algorithm is then scrutinized with respect to its feasibility for distributed implementation. In our future work, new methods of relaxing the fixed route assumption and the interference graph model will be investigated and also a proof-of- concept implementation of the propOsed distributed algorithm with the long-term reward mechanism will be implemented.

REFERENCES

[1] Dongyue Xue and Eylem Ekici, "Cross- Layer Scheduling for Cooperative Multi Hop Cognitive Radio Networks", *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 31, NO. 3*,MARCH 2013, 0733-8716/13/ IEEE

[2] I.Akyildiz, W. Lee, M. Vuran, and S. Mohanty,"NeXtgeneration/ dynamic spectrum access/cognitive radio wireless networks: *A survey, Computer Networks", in Computer Networks J. (Elsevier), vol. 50,* no.13, pp. 2127-2159, 2006.

[3] L. Lai, H. Gamal, H. Jiang and H. Poor, *"Cognitive Medium Access: Exploration, Exploitation and Competition", in IEEE Trans. Mobile Computing, vol. 10,* no. 2, pp. 239-253, Feb. 2010.

[4] S. Bayat, R. Louie, Y. Li, and B. Vucetic, *"Cognitive radio relay networks with multiple Primary and secondary users: distributed stable matching algorithms for spectrum access", inProc. IEEE International Conference on Communications (ICC) 2011*, pp. 1-6, Jun2011.

[5] P. Huang, X. Lin, and C. Wang, *"A low- complexity congestion control and scheduling algorithm for multi-hop wireless networks with order optimal per-flow delay", in Proc. IEEE INFOCOM'11,* pp. 2588-2596, April 2011.

[6] Y. Han, A. Pandharipande, and S. H. Ting, "*Cooperative decode and forward relaying for secondary spectrum access", in IEEE Trans.Wireless Commun., vol 8, no. 10*, pp. 4945-4950, Oct.

LI-FI BASED AUDO COMMUNICATION AND DEVICE SWITCHING

Madan Kumar K, Karthik M C, Manoj Kumar C, Harshith C,D Pradeepa,K R Pavan Kumar
Department of Electrical And Electronics ,Visveswaraya Technological University, Bangalore-60

madankumar0000031@gmail.com,karthikmc007@gmail.com,manojkumar5651488@gmail.com,hrsthrocks@gmail.com

ABSTRACT-From many years, there is a rapid development in the solid state LED materials which gave way for the future generation data communication known as visible light communication(VLC). VLC has a promising future and it acts as a complement to the RF communication by achieving larger bandwidth and high data rate. At present, the day to day activities use lot of LED based lights. which can also be used for communication because of the advantages like fast switching, high efficiency and safe to human vision. Hence, this paper presents about audio communication and device switching through visible light which consists of the LEDs that transmit data to the receiver. The receiver circuit consists of solar panel or power led based on its application connected with the amplifier and output circuit to recover back the amplified version of original input signal.

Keywords: Light Feudality, Wireless Feudality, RF Communications.

I. INTRODUCTION

One of the most evolving technologies in industry today is the wireless technology. In wifi technology it has many drawbacks like RF waves can harm living beings which may cause many skin diseases where new wireless technology came in to existence called LI_FI which will overcome the drawbacks of WI-FI because it uses visible light for communication which is harmless. As we know, speed of light is $3*10^8$ m/s so that the problem of speed in WI-FI can be relieved .it also has many applications like device switching audio communication and data transfer.

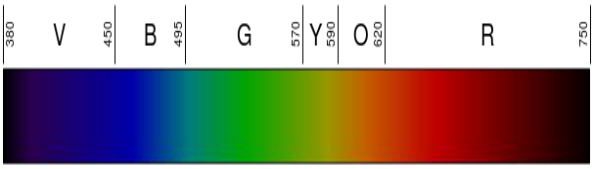


Fig.1.2 The visible spectrum

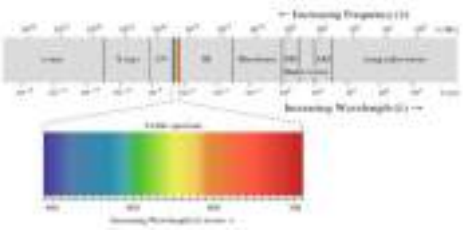


Fig.1.1 The Electromagnetic Spectrum

1.1 OBJECTIVE

To design and implement a communication system in the real time environment using Li-Fi, such that it can communicate with maximum speed and with better efficiency. And also to build a Switching system for commercial use in home\office appliances which can be switched ON and OFF very quickly, for real time application.

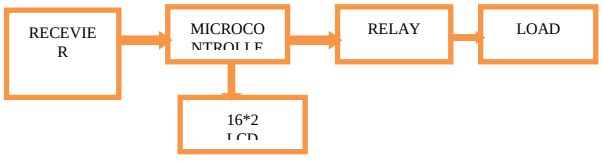
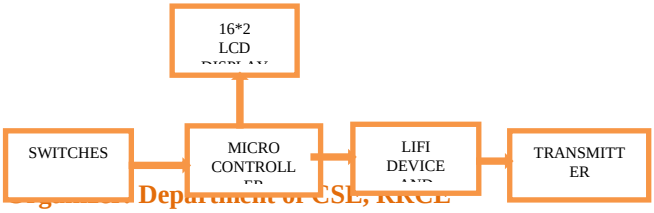


Fig 1.1 shows the electromagnetic spectrum of different frequency and wavelength for communication medium. Fig 1.2 shows the wavelength associated for visible light medium. Visible light communication (VLC) data communication medium using visible light operates between 400 THz (780 nm) and 800 THz (375 nm). Li-Fi transmits data using the spectrum of vlc at a speed of 10 to 20Gbit/s more than 200 times faster than 'superfast' data device [12].

II. BLOCKDIAGRAM AND METHODOLOGY



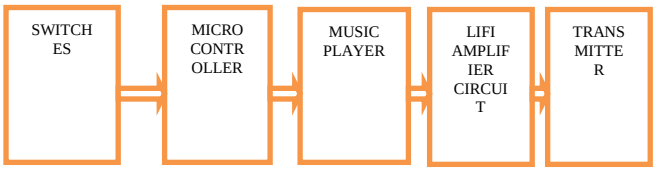


Fig 2.1 Block diagram of transmitter

From all the block diagrams the blocks are explained as follows

SWITCHES

- the lifi module which consists of switches (push buttons) is connected to microcontroller with different ports
- this same controller is connected to the 16*2 lcd display, which notify the on of state of switches and displays on lcd.

LIFI DEVICE

- the lifi circuit consisting of coil,led light and a battery
- the input is fed through the controller and this high frequency low voltage is amplified through the amplifier.
- Due to the current modulation the intensity of light varies propositionally .through this illumination data is transferred, with output high.

PHOTO TRANSISTOR

- It acts as receiver for device switching module.
- Initially the output of photo transistor will be zero,even if the supply is given because base voltage is zero.
- When the base voltage is applied from the transmitter through led ,the output of photo transistor will be high ,hence current flows through the diverted path.

RELAY

- It senses the output of microcontrollerand it tripsto give output to the load depends on its rating[type =PCB relay].

LOAD

- Any type of load can be used either AC or DC load.

SOLAR PANEL

- It is a type of photo cell ,which receives light illumination falling on it and this is converted in to electrical signal.

SPEAKER

- It is the output device which converts electrical signal to original sound signal.

2.2 ADVANTAGES:

- It is free from bandwidth.
- very low maintenance cost.
- Very Cheaper than Wireless-Fidelity.
- Theoretical speed up to 1 GB per second, higher speed than wifi .
- Lower electricity costs.
- Vlc is for free and it will not harm human vision

III. IMPLEMENTATION

Implementation results and micro controller kit figures 2.5, 2.6, 2.7 are shown below:

Fig 2.5 Device switching

Fig 2.6 Microcontroller with LCD

Fig 2.7 Audio receiver with solar panel

IV. APPLICATIONS

1. SMART LIGHTING

Smart buildings like MNC's, hospitals, theaters etc require smart lighting. for smart communication where lifi provides illumination as well as secured communication.

2. HAZARDOUS ENVIRONMENTS

Wifi can't be used in hazardous environments like petro chemical plants, mining plants due to chemical reactions maytakesplace, lifi provides visible light which will not harm.

3. OTHER APPLICATIONS

- Mobile connectivity, vehicle and transportation, defense and security, hospitals and health care etc

V. CONCLUSION

1) *The Wi-Fi technology which uses radio frequency to broadcast the data and currently offers very high data rate but insufficient for moving large data files like HDTV movies, music libraries and video games.The use of radiofrequency communication devices is increasing widespread, some emissions and intensities interfere with sensitive electronic equipment brings severe damages. To remedy this limitation of current technology, we can utilize visible light communication [VLC], known as Li-Fi technology. Li-Fi appears to be an important potential component in expanding usable bandwidth, protecting sensitive electrical equipment and data, creating more biologically friendly communications technology, and helping to develop seamless computing applications.*

REFERENCES

[1] DHAKANEVIKASNIVRUTTI, RAVI RAMCHANDRANIMBALKAR, "LIGHT-FIDELITY: A RECONNAISSANC OF FUTURE TECHNOLOGY"- VOLUME 3, ISSUE 11, NOVEMBER 2013

[2] Tsonev, D.;Sinanovic, S.; Haas, Harald (15 September 2013). An IEEE Standard for Visible Light Communicationsvisiblelightcomm.com, dated April 2011.. "Complete Modeling of Nonlinear Distortion in OFDM-Based Optical Wireless Communication". *IEEE Journal of Lightwave Technology* 31 (18): 3064–3076.doi:10.1109/JLT.2013.2278675

[3] Tony Smith (24 May 2012). "WTF is... Li-Fi? Optical data transfer's new leading light?".*The Register*. Retrieved 22 October 2013.

[4] Iain Thomson (18 October 2013). "Forget Wi-Fi, boffins' get 150Mbps Li-Fi connection from a light bulb: Many (Chinese) hands make light work".*The Register*. Retrieved 22 October 2013.

[5] "Visible-light communication: Tripping the light fantastic: A fast and cheap optical version of Wi-Fi iscoming".*The Economist*. 28 January 2012. Retrieved 22 October 2013.

[6] Technopits.blogspot.comtechnology.cgap.org /2012/01/11/a-Li-Fi-world/

[7] Haas, Harald (July 2011)."Wireless data from every light bulb".*TED Global*. Edinburgh, Scotland.

[8] T. Ishii , M. Hirose and H. Iwata "Automatic Recognition of Driver\'s Facial Expression by Image Analysis", Journal of JSAE, vol. 41, no. 12, pp.1398 -1403

[9] M. Kaneda "Development of a Drowsiness Warning System", The 14th International

[10] www.lificonsortium.org

[11] http://beyondweblogs.com/what-is-Li-Fi-is-this-replacing- Wi-Fi/

[12] http://en.wikipedia.org/wiki/Li-Fi

[13] Electronic Communication Systems – George Kennedy

[14] Electronics in Industry - George M.Chute

[15] Principles of Electronics - V.K.M

Wireless robot control with robotic arm using mems and zigbee

RAKESH. H .K, SHIVASHANKARA. B .S.

Department of Mechanical engg., MTech Student (IAR), VTU, MCE, Hassan-573201
Department of Mechanical engg., Asst. professor (IAR), MCE, Hassan-573201
Email: rakeshhk26@gmail.com

ABSTRACT- In numerous use of controlling automated contraption it turns out to be very hard and muddled when there comes the piece of controlling it with remote or various switches. A Gesture Controlled robot is a sort of robot which can be controlled by hand motions not by old catches. Simply need to wear a little transmitting gadget close by which incorporated a quickening meter. This will transmit a suitable order to the robot with the goal that it can do whatever we need. The transmitting gadget incorporated an ADC for simple to computerized transformation and a microcontroller IC which is use to encode the four piece information and after that it will transmit by a ZigBee Transmitter module. At the less than desirable end ZigBee Receiver modules gets the encoded information and translate it by microcontroller IC. This information is then prepared by a microcontroller lastly engine driver attempts to control the engines. By including deterrent sensor and temperature sensor robot is improved to work in various circumstance and environment.

Keywords - ZigBee, Embedded, MEMs, Gesture, Robotics

I. INTRODUCTION

The presently human-machine connection is moving far from keypad and is turning out to be a great deal more perfect with the physical world. With every passing day the hole amongst machines and people is being lessened with the acquaintance of new advancements with facilitate the way of life. Innovation has assumed an extremely noteworthy part in enhancing the personal satisfaction. One route through which this is done is via mechanizing a few errands utilizing complex rationale to streamline the work. Signals have assumed a key part in lessening this pit. Presently a day's robots are controlled by remote or mobile phone or by direct wired association. Contemplating cost and required equipment's this things builds the multifaceted nature, particularly for low level application. For instance, in telerobotics, slave robots have been shown to take after the expert's hand movements remotely.

Robots are turning out to be progressively valuable on the war zone since they can be furnished and sent into perilous regions to perform basic missions. Controlling robots utilizing customary techniques may not be conceivable amid clandestine or unsafe missions. MEMs based robot was produced for interchanges in these great situations where writing on a console is either illogical or unimaginable. This paper reports an adjustment of this correspondences for transmitting motions to a military robot to control its capacities. Signals were utilized to control a pick and place robot model.

This framework can explore the remote robot in the diverse environment utilizing different motions orders. In this framework, the robot works from transmitter area with a decent quality in-constructed accelerometer sensor. Mems sensor is utilized to peruse the position of hand motions to produce summons for the robot and is taken as an info, position of Mems is then used to extricate the signal order. The summon prepared by expert microcontroller and is sent through the ZigBee transmitter. At ZigBee recipient area gets information from slave microcontroller through ZigBee module and as indicated by Mems position; information got the collector side microcontroller prepared to incite the engine by driver circuit associated with ARM7microcontroller. At long last the robot is moved in every single conceivable bearing in the earth utilizing six conceivable sorts of charges which are Forward, Backward, Right, Left and arm developments. Signal charge can have one of the six conceivable orders as indicated.

II. PREAMBLE

AIM: The primary point of this anticipate is give the material taking care of through robot without association of people, rather robot is made remote and controlled through expert board utilizing motion position and ZigBee module.

OBJECTIVE & SCOPE: This anticipates has a decent degree in the field of route which gives the security to people where they can't reach through unsafe spot. This framework is exceptionally financially savvy to supplant other previous framework. Have proposed the framework which is helpful for some reasons like military reason and local reason to keep away from human danger .It is simple and easy to use approach to control robot. This framework gives easy to understand control over robot.

III. EXISTING SYSTEM

Numerous frameworks exist that are utilized for controlling the robot through motions. Some signal acknowledgment

frameworks include, versatile shading division , hand finding and marking with blocking, morphological sifting, and afterward motion activities are found by format coordinating and skeletonising. This doesn't give dynamicity to the motion inputs because of layout coordinating. Another framework utilizes machine interface gadget to give continuous signals to the robot. Simple flex sensors are utilized on the hand glove to quantify the finger twisting , likewise hand position and introduction are measured by ultrasonics for signal acknowledgment . The rise of administration robots in mid 90's trailed by the advancement of Natural dialect interface through console has been given by Torrance in 1994.

In 2008, Chinese movement police framework utilized two 3-hub accelerometers altered on the back of their arms that were synchronized with activity lights. In 2010, Sauvik Das et al have utilized an accelerometer as a potential spying gadget to show areas and exercises of client without one's information. One of the restrictions was that inbuilt accelerometer Smartphone would need to be in the same spot as was in the preparation mode to make exact expectations. In 2010, Smartphone's were utilized to control Universal Robot Control System by the understudies of Kyungpook National University, Korea, to plan an ongoing robot control framework in omnipresent environment. Specialist's proposed vision-based interface that included signal acknowledgment through camera to give geometrical data to the robots. They created versatile robot frameworks that were told through arm position.

IV. SYSTEM DESIGN MODEL

A. SOFTWARE DESIGNING:

For the operation reason, the client application directions are composed programming code by utilizing implanted c. The application system is accumulated by utilizing KEIL-C compiler and proselytes the source record into .hex document. Project is dumped by the utilization of small scale streak software engineer. Here the system is dumped in the microcontroller ROM memory area. The µVision4 screen furnishes us with a menu bar for charge passage, a device bar where we can quickly choose order catches, and windows for source records, exchange boxes, and data shows. µVision4 lets us all the while open and view different source documents.

An undertaking contains enough data to take an arrangement of source documents and produce precisely the paired code required for the application. In light of the high level of adaptability required from the devices, there are numerous choices that can be set to arrange the instruments to work in a particular way. It is repetitive to need to set these choices up each time the application is being fabricated; in this manner they are put away in a task record. Stacking the venture document into KEIL illuminates KEIL which source records are required, where they are, and how to design the instruments in the right way. The client of KEIL focuses on "tasks". An undertaking is a rundown of all the source records required to manufacture a solitary application, all the apparatus alternatives which determine precisely how to assemble the application, and – if required – how the

application ought to be reproduced. The task can then be spared to safeguard the settings. The venture is reloaded and the test system or debugger began, all the fancied windows are opened. KEIL venture documents have the expansion.

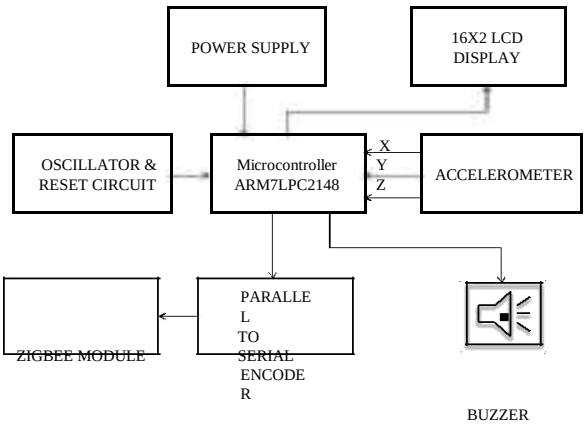
B. HARDWARE DESIGN MODULE

In cases there is a requirement of a mechanism where robot should be controlled without any physical contact. Therefore gesture is choice in order to achieve this primary goal. Accelerometer sensor that takes gesture as its input can do this job. The Power supply is a reference to a source of electrical power. The ADXL335 is a small, thin, low power, complete 3-axis accelerometer with signal conditioned voltage outputs. The product measures acceleration with a minimum full-scale range of ±3 g. It can measure the static acceleration of gravity in tilt-sensing applications, as well as dynamic acceleration resulting from motion, shock, or vibration. ZigBee is the name of a specification for a suite of high level communication protocols using small, low-power digital radios based on In cases there is a prerequisite of an instrument where robot ought to be controlled with no physical contact. Accordingly signal is decision with a specific end goal to accomplish this essential objective. Accelerometer sensor that takes motion as its information can carry out this occupation. The Power supply is a reference to a wellspring of electrical force. The ADXL335 is a little, thin, low power, complete 3-pivot accelerometer with sign adapted voltage yields. The item measures increasing speed with a base full-scale scope of ±3 g. It can gauge the static quickening of gravity in tilt-detecting applications, and in addition dynamic speeding up coming about because of movement, stun, or vibration. ZigBee is the name of a detail for a suite of abnormal state correspondence conventions utilizing little, low-control advanced radios in light of the IEEE 802.15.4-2006 standard for remote individual zone systems (WPANs, for example, remote earphones associating with mobile phones by means of short-range radio. The innovation is planned to be easier and less expensive than different WPANs, for example, Bluetooth. ZigBee is focused at radio-recurrence (RF) applications that require a low information rate, long battery life, and secure systems administration. ZigBee works in the modern, investigative and restorative (ISM) radio groups; 868 MHz in Europe, 915 MHz in nations, for example, USA and Australia, and 2.4 GHz in many locales around the world. The innovation is expected to be less complex and less expensive than different WPANs, for example, Bluetooth. ZigBee chip merchants ordinarily offer coordinated radios and microcontrollers with somewhere around 60K and 128K blaze memory, for example, the Free scale MC13213, the Ember EM250 and the Texas Instruments CC2430. Radios are likewise accessible stand-alone to be utilized with any processor or microcontroller. For the most part, the chip merchants additionally offer the ZigBee programming stack, albeit autonomous ones are likewise accessible. ZigBee gadgets are required to fit in with the IEEE 802.15.4-2003 Low-Rate WPAN standard. The standard determines the lower convention layers—the physical layer (PHY), and the medium access control (MAC) segment of the information join layer (DLL). This standard determines operation in the unlicensed 2.4 GHz, 915 MHz and 868 MHz

ISM groups. In the 2.4 GHz band there are 16 ZigBee channels, with every channel requiring 5 MHz of transmission capacity. The inside

recurrence for every channel can be ascertained as, FC = (2350 + (5 * ch)) MHz, where ch = 11, 12... 26. The essential channel access mode is "bearer sense, various access/impact evasion" (CSMA/CA). That is, the hubs talk similarly that individuals chat; they quickly verify that nobody is talking before they begin. Signals are sent on an altered planning plan, and don't utilize CSMA. Message affirmations likewise don't utilize CSMA. At long last, gadgets in Beacon Oriented systems that have low idleness ongoing necessities may likewise utilize Guaranteed Time Slots (GTS), which by definition don't utilize CSMA. These ICs are matched with each other. For legitimate operation a couple of encoder/decoder with the same number of location and information organization ought to be chosen. The Decoder get the serial location and information from its relating decoder, transmitted by a transporter utilizing a ZigBee transmission medium and offers yield to the yield pins subsequent to preparing the information. The L293 and L293D are fourfold high-ebb and flow half-H drivers. The L293 is intended to give bidirectional drive streams of up to 1 An at voltages from 4.5 V to 36 V. The L293D is intended to give bidirectional drive streams of up to 600-mA at voltages from 4.5 V to 36 V. Both gadgets are intended to drive inductive loads, for example, transfers, solenoids, dc and bipolar venturing engines, and in addition other high present/high-voltage loads in positive-supply applications.

Transmitting section: The beneath transmitting graph shows the transmitting segment which incorporates an accelerometer whose yield is in persistent structure as the encoder can just comprehends the computerized information and this information is to be transmitted utilizing zigbee module which transmits the serial information changed over by the encoder from parallel information.



composed the calculation here, to show the nearness of a hindrance. The temperature sensor yield sign is sustained to the microcontroller in which an appropriate installed "c" project is composed the calculation here, to show the temperature of the earth around robot. The metal sensor yield sign is sustained to the microcontroller in which an appropriate implanted "c" project is composed the calculation here, to show the nearness of a metal which incorporates a sounding of bell that speaks to a bomb.

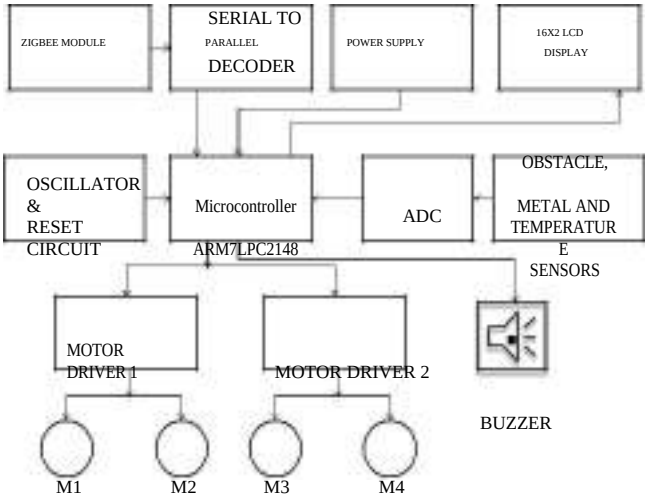


Figure 2. Receiver section

V. RELATED WORKS

Figure 1. Transmitter section

Receiving Section: The underneath getting square chart demonstrates the recipient segment the transmitted information by the transmitter is gotten by the zigbee module and the serial information is given as contribution to the decoder which changes over the serial information to parallel information and is given as contribution to the microcontroller which comprises of a predefined project to satisfy our assignment, contingent on the information got the controller produces some signs to the engine driver, LCD, ringer, and so forth., here the reason for the engine driver is to drive the engines and here LCD and bell are utilized for some particular signs. The infrared sensor yield sign is sustained to the microcontroller in which a reasonable inserted "c" project is

Working of robot:

a. ROBOT MODE:

Based on hand movements with holding accelerometer inside, changing the movement robot works as follows: In forward: ROBOT moves in forward direction.

In reverse: ROBOT moves in reverse direction.
In left: ROBOT moves in left direction

In right: ROBOT moves in right direction

b. ROBOT ARM MODE:

Based on hand movements with holding accelerometer inside, changing the movement robot works as follows: In forward: ROBOT ARM opens up.

In reverse: ROBOT ARM closes. In
left: ROBOT ARM slides down. In
right: ROBOT ARM slides up.

c. ROBOT WITH SENSORS

Sensor is a refined gadget whose capacity is to recognize and measure any non-electrical parameters, for example, temperature, weight, dampness, speed, weight and so on and change over it into a sign which can be measured electrically for e.g. voltage.

For the usage reason sensor ought to be precise, not reliant on ecological conditions, with extensive variety of qualities and high determination, direct, financially savvy and exceptionally aligned. So for this reason for the most part IC sensors are utilized. In our project we have mainly implemented four sensors:

K. Metal sensor: whenever metal is detected the robot is stopped for some delay, say 5s.

L. Temperature sensor: Temperature is set to know environment of robot. Whenever the temperature reaches above the limited temperature, the robot stops.

M. Two IR sensors: this sensors used to detect the obstacle which comes infront while robot is moving forward and to detect the obstacle which comes behind while robot is moving reverse direction.

VI. EXPERIMENTAL RESULTS

Figure 3. TRANSMITTER MODULE

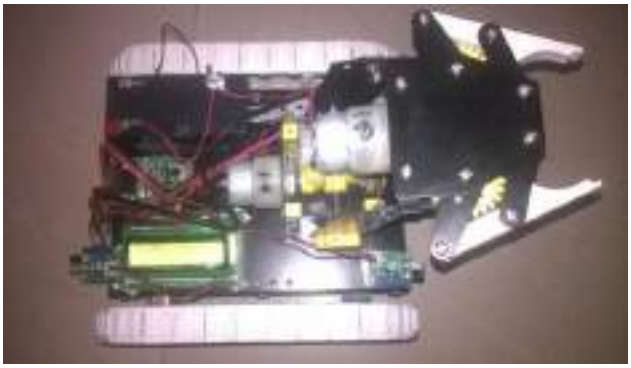
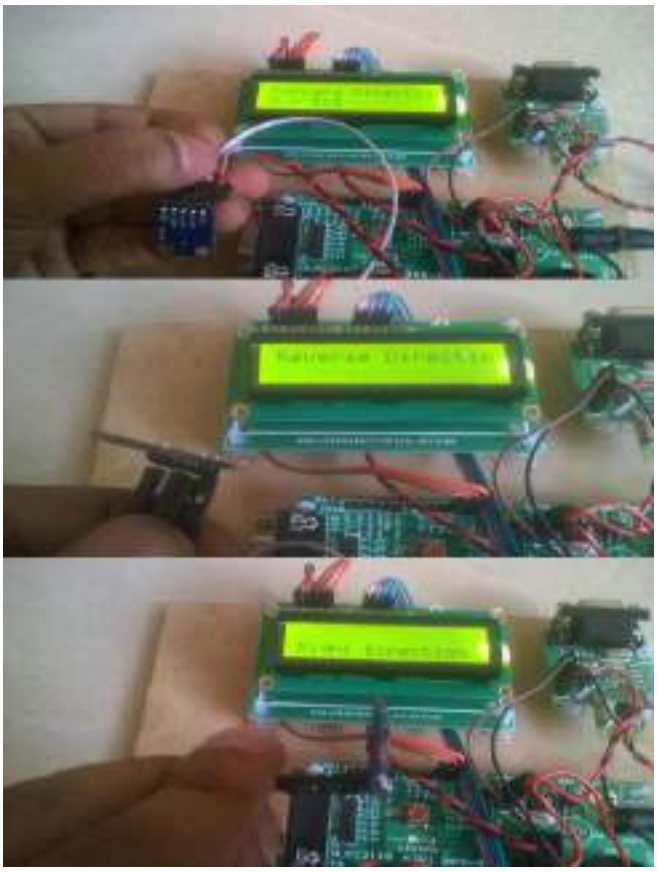


Figure 4. RECIEVER MODULE

4



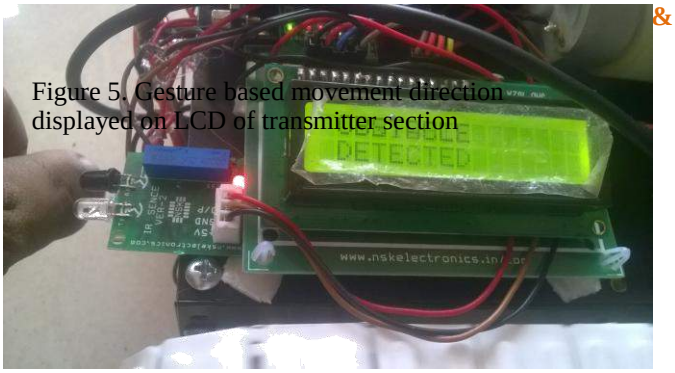


Figure 5. Gesture based movement direction displayed on LCD of transmitter section

Figure 6. Gesture based movement of robotic arm

Figure 7. Obstacle detection by IR sensor

VII. FUTURE SCOPE:

As of now control the robot is with signal position .In future can build up the task to Control vehicle. By improvement can build up the task to control the remote robot by two way correspondence which prompts robot cooperation.

VIII. CONCLUSION:

In our arrangement of motion controlled robots, we have just considered a predetermined number of signals. Our calculation can be stretched out in various approaches to perceive a more extensive arrangement of motions. The motion acknowledgment bit of our calculation is excessively basic, this strategy would require, making it impossible to be utilized as a part of testing working conditions. This kind of control could enhance efficiency, decrease the impacts of dull movements, and enhance security. Progressed mechanical arms that are outlined like the human hand itself can undoubtedly controlled utilizing hand signals as it were. The mechanical arm will copy the development of the controller. Progressed mechanical arms like these can perform unpredictable and perilous errands easily. Proposed utility in fields of development, perilous waste transfer, and therapeutic sciences.

IX. REFERENCES

KK. T. H. Speeter (1992), "Transformation human hand motion for tele manipulation," *Presence*, **1**, 1, pp. 63–79.

LL. S. Zhou, Z. Dong, W. J. Li, and C. P. Kwong (2008), "Hand-written character recognition using MEMS motion sensing technology," in *Proc. IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics*, pp.1418–1423.

W. T. Freeman and C. D. Weissman (1995) , "TV control by hand gestures, "presented at the IEEE Int. Workshop on

Automatic Face and Gesture Recognition, Zurich, Switzerland.

D. Xu (2006), "A neural network approach for hand gesture recognition in virtual reality driving training system of

SPG," presented at the *18th Int. Conf. Pattern Recognition*.

[9] H. Je, J. Kim, and D. Kim (2007), "Hand gesture recognition to understand musical conducting action," presented at the *IEEE Int. Conf. Robot &Human Interactive Communication.*

[6] T. Yang, Y. Xu, and A. (1994) , Hidden Markov Model for Gesture Recognition,CMU-RI-TR- 94 10, Robotics Institute, Carnegie Mellon Univ.,Pittsburgh, PA.

[7] S. Zhou, Q. Shan, F. Fei, W. J. Li, C. P. Kwong, and C. K. Wu *et al* (2009)., "Gesture recognition for interactive controllers using MEMS motion sensors," in *Proc. IEEE Int. Conf. Nano /Micro Engineered and Molecular Systems*, pp. 935–940.

[8] S. Zhang, C. Yuan, and V. Zhang (2008), "Handwritten character recognition using orientation quantization based on 3-D accelerometer," presented at the *5th Annu. Int. Conf. Ubiquitous Systems.*

[9] J. S. Lipscomb (1991), "A trainable gesture recognizer,"

Pattern. Recognit.,**24**, 9, pp. 895–907.

[10] D. H. Rubine (1991), "The Automatic Recognition of Gesture," Ph.D dissertation, Computer Science Dept., Carnegie Mellon Univ., Pittsburgh, PA.

An Autonomous Robot for Finding Optimal Path during Dynamic Environment Using Grid Based Approach

JASHWANTH N B, SREE RAJENDRA
Department of Mechanical engg., MTech Student (IAR), VTU, MCE, Hassan
jashu5457@gmail.com,sreerajendra@gmail.com

ABSTRACT- The Proposed framework is to give a way choice calculation to versatile robots in extensive element/obscure situations. The proposed calculation permits a portable robot to explore through static hindrances, and finding the way keeping in mind the end goal to achieve the objective without crash. The point of the work is to diminish the danger of impacts and time of way following in situations when the robot over and again navigates between predefined target focuses (e.g. for transportation or review errands). The calculation is usable regardless of the fact that almost no is thought about the earth or on the off chance that it gets totally rebuilt amid the mission. We are pointed in demonstrating our test results with our proposed way choice calculation that will have the capacity to lessen impact hazard, travel time and travel separations in huge element situations.

Keywords -Dynamic Environment, Navigation, Obstacles, Optimal Path, Robot

I. INTRODUCTION

Artificial Intelligence (AI) is the knowledge showed by machines or programming. It is furthermore the name of the scholastic field of study which thinks how to make PCs and PC programming that are fit for cunning behavior.

The focal issues (or objectives) of AI exploration incorporate thinking, information, arranging, learning, characteristic dialect handling (correspondence), observation and the capacity to move and control objects. General insight is still among the field's long haul objectives. As of now mainstream approaches incorporate factual strategies, computational insight and conventional typical AI. There are a substantial number of instruments utilized as a part of AI, including adaptations of inquiry and scientific streamlining, rationale, strategies taking into account likelihood and financial matters, and numerous others. The AI field is interdisciplinary, in which various sciences and callings merge, including software engineering, arithmetic, brain science, etymology, theory and neuroscience, as they as other particular fields, for example, simulated brain science.

The proposed system is to provide a path selection algorithm for mobile robots in the large dynamic environments / unknown. The proposed algorithm allows a mobile robot to navigate through the static obstacles, and find the way to reach the target without collision. The goal of the work is to reduce the risk of collisions and the time following routes where the robot traverses repeatedly between the preset target points (eg for transportation or inspection tasks). The algorithm is usable even if very little is known about the environment or if it is completely restructured during the mission. They are intended to show our test results with our path selection algorithm project that will be able to reduce the risk of collision, the travel time and travel distances in large dynamic environments.

A robot is a machine designed to perform one or more tasks repeatedly, with speed and accuracy. There are so many different types of robots as there are tasks they perform. Robotics is the branch of computer science that deals with the design, construction, operation and application of robots, as they used as computer systems for control, sensory feedback, and information processing. These technologies dealing with automated machines that can take the place of man in hazardous environments or manufacturing processes , or resemble humans in appearance, behavior and cognition or . Many robots today are inspired by nature contributing to the field of bio-inspired robotics

II. EXISTING SYSTEM

Through framework clients is allocated to draw different courses between a source and the destination. At long last when our applications are begun, it forms every course and changes over it into robot justifiable directions and helps our application to controls the robot over a serial connection with any remote innovation for robot correspondence. The line devotee robot they have arranged which utilizes line directions through framework will maintain a strategic distance from the overhead of drawing a physical line on the floor in such vast environment which is supporter as the present idea in numerous line adherent robots.

Fig1:Line Follower Robot

III. PROPOSED SYSTEM

The approach they move forward is motivated by the fact that our robot is made repeated traversal between the preset target points (from source to destination through. Various Paths) in a dynamically changing environment. Examples of such implementations are fetch-and-carry task of industrial and agricultural applications or visiting certain checkpoints in applications for security and surveillance. Efficient operation of robot should fulfill its mission as quickly and as safely as possible. This means that it is worth to avoid situations where the robot is forced to re-plan your itinerary, make a detour, can lead into a dead end or collide with unexpected obstacles by modeling the environment or learning of its properties, delays can be minimized and the risk can be reduced and helps to choose a path that is easy to follow, which if free of obstacles. In a dynamic environment with an unknown obstacle to distribution, the best path to the goal is not necessarily the shortest. Depending on the nature of the environment, there may be roads that are longer but easier to follow. By introducing a path generation algorithm, the robot can test several predefined alternatives to achieve the goal. Remembering his experiences following way, he can learn to follow the paths that save time and reduce risk. As the environment changes, the robot will re-evaluate its past experience and to adapt to using new roads easily passable.

Dynamic Obstacle handling - sensors will be installed in the front of the robot. When an unknown obstacle appears in the following location on the line, it is detected by the sensor. All the distance covered by the robot will be displayed on the PC screen.

Fig 2: Autonomous Robot

IV. ALGORITHM

BFS (Breadth First Search) is a calculation that is utilized to cross the way from source to destination utilizing the most limited way. As the name itself says it first visits the broadness and after that goes to its comparing profundity, thusly it crosses to its destination utilizing the most limited way

HOW IT WORKS:

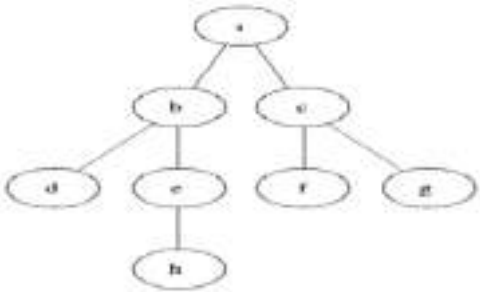


Fig 3. A tree structure.

PSEUDOCODE

Input: A graph Graph and a starting vertex rootof Graph
Output: All vertices reachable from root labeled as explored.
A non-recursive implementation of breadth-first search:
Breadth-First-Search(Graph, root)
For each node n in Graph;
n.distance = INFINITY
n.parent = NIL
create empty queue Q
root.distance = 0
Q.enqueue(root)
while Q is not empty;
current = Q.dequeue()
for each node n that is adjacent to current:
if n.distance == INFINITY;
n.distance = current.distance + 1
n.parent = current
Q .enqueue(n)

Moore Neighborhood counter algorithm

Moore Neighborhood of a pixel, P, is the arrangement of 8 pixels which impart a vertex or an edge to that pixel. The fundamental thought is: - When the present pixel p is dark, the Moore neighborhood of p is analyzed in clockwise heading beginning with the pixel from which p was entered and propelling pixel by pixel until another dark pixel in P is experienced. The calculation ends when the begin pixel is gone to for second time. The dark pixel strolled over will be the shape of the example.

.**ALGORITHM** Modified Moore's neighbor algorithm
Input: A square tessellation T containing a connected component P of black cells.
Output: A sequence B (b1, b2… Bk) of boundary pixels i.e. the contour line. We define M (p) to be the Moore neighborhood of pixel p, c denotes the current pixel under consideration i.e. c is in
M(p).
Begin
Set B to be empty.
From bottom to top and left to right scan the cells of T until a black pixel, s, of P is found.
Insert s in B.
Set the current boundary point, p, to s i.e. p = s.
Set c to be the next clockwise pixel in M(p).
While c is not in B do

If c is black
Insert c in B.
Set p =c.
End if
Advance c to the next clockwise pixel in M (p).
End while
End
In handling the portrayal, we separate what is thought to be the "principle bearing" of the item as for the robot and discrete it into one of 8 conceivable headings.

V.IMPLEMENTATION

The proposed system works on the basis of 5 modules they are namely Grid formulation modulation module, location and wall marking module, path finding module, command extraction module, command transmission module and last two modules works using Moore's neighbor contour tracing algorithm and as explained above it works based on grid based approach.

i. **Grid formulation module**: Create grid cells based on the given no. of rows and columns. This module uses graphic object to draw rectangles of cells in a control panel using an object colored pencil. The whole workspace is divided into grids. It creates grid cells based on the given no. of rows and columns. Here we are taking 5*5 grids as reference and constant cell size of 90 pixel, depending on the given workspace environment, the cell-size can be increased which proportionally increases workspace. The index always begins with zero(0).Since it is a 5*5 matrix, 0th index starts from bottom right and ends with index four(4) and (0,0) is always the default source index.
This module uses graphics object to draw cell rectangles over a panel control using a colored pen object.

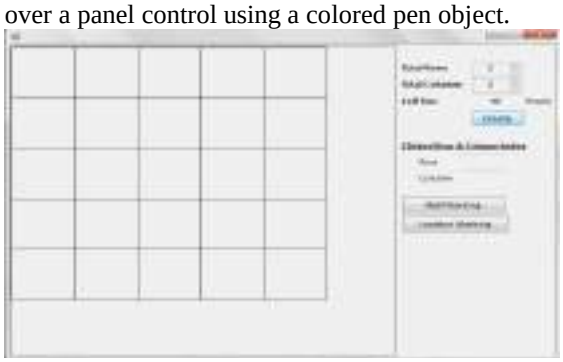


Fig 4:Grid formulation module

ii.Location and Wall marking module: This module uses graphic object to draw solid rectangles over a grid cell using an object selected brush color. This module uses different colored brushes for wall and marking place. This module uses graphics object to draw solid rectangles over a selected grid cell using a colored brush object. This module uses different colored brushes for wall and location marking. As shown in the below screenshot, we have marked the static obstacle using grey color. These grid are indicated in the pattern like (2, 4) (2, 2) (2, 1) (2, 0). The destination is marked using dark blue color. Index representing destination are (0, 2) and (0, 1). Multiple destinations can also be marked.

Fig 5: Location and Wall marking module

iii.**Path Finding Module**: This module helps you find the shortest path between the current position of the robot and a destination location algorithm using breadth-first search (BFS). This module is also responsible for recalculating alternate routes between the source and destination obstacle detection for dynamic.

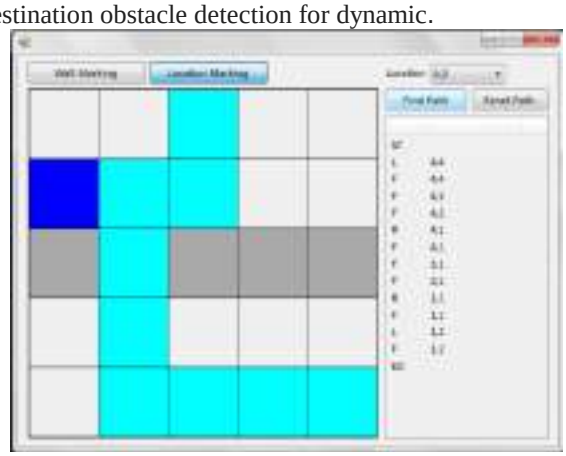


Fig 6: Path finding module

As shown in the screen shot above here the destination is selected as the (0, 2) and if we click the find path button the best feasible path from the source (0, 0) and destination (4, 2) is calculated and it is displayed by the light blue color. Clicking on the find path button will also list the instruction that has to be followed by robot. For example in this case, the robot will be in the upright position, the 1st command is 'L' then it turn to the left move to the breadth of that particular grid using **BFS** algorithm. The next instruction is 'F' and robot traverse to the forward direction and continues with commands generated **MVC** algorithm

iv.Command Extraction module(Using Moore Neighbor contour tracing algorithm) :This helps in converting the shortest route module is between the source and the destination location in understandable by our robot commands for navigation. Here commands like "L" will be computed for the manufacture of robot to move left, "R" to move right, "F" to advance and "B" to go back.

v.Command Transmission / Receive module: This module is used for transmitting commands to a robot via wireless RF / ZigBee transceiver connected to the COM port on our system. This module will also be responsible for dynamic obstacles receive alerts sent by the remote robot during its movement to the selected destination.

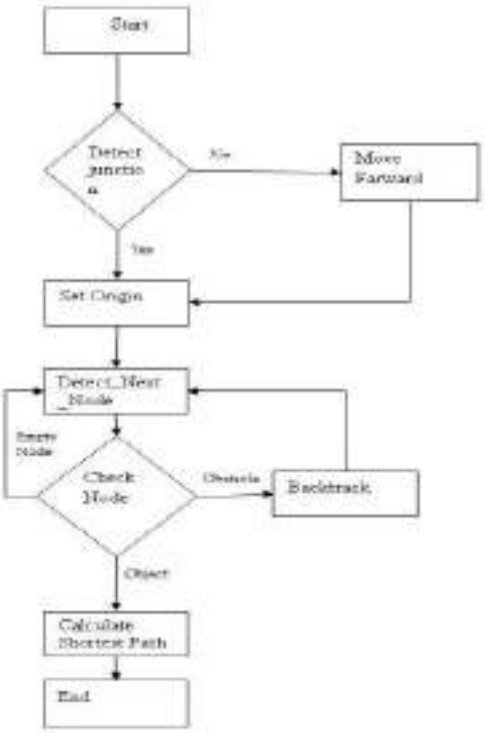


Fig 7: Implementation flow chart.

After setting the destination point the robot will be in the source node it first checks for the empty grids to move, if there is any obstacle in in the neighbor node it checks other child nodes using breadth first search algorithm based on that it will calculate optimal/shortest path.

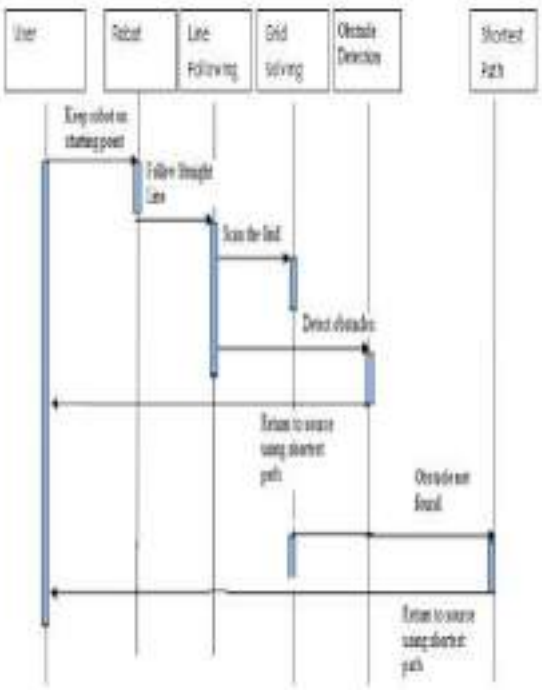


Fig 8: sequence diagram

After destination is set the robot will be in source position it will follow the straight line for empty grids using breadth first search algorithm and if there is any dynamic obstacles it will backtrack to source position using optimal path if there is no obstacles it will reach destination using shortest path

VI. APPLICATION

Case of this sort of usage is to bring and-convey assignment of mechanical and farming applications or going by specific checkpoints in security and reconnaissance applications. A proficiently working robot is required to satisfy its task as quick and as securely as could reasonably be expected. It implies that it is advantageous to maintain a strategic distance from circumstances where the robot is compelled to re-arrange its course, take a makeshift route, can crash into a halt or slam into surprising deterrents And How they were known by demonstrating nature or taking in its properties, the time postponements can be minimized and the danger can be decreased and helps picking a way that is anything but difficult to take after which if free from impediments. In a dynamic situation with an obscure hindrance circulation, the best way to the objective is not as a matter of course the briefest one. Contingent upon the way of nature, there may exist courses that are longer yet less demanding to take after.

1.Transportation: they utilize Mobile Robots in healing centers for transportation reason to convey examples, pharmaceuticals and dinners and so on.

2.Inspection tasks: Autonomous mechanical technology would spare and safeguard human life by evacuating serving officers who may somehow be executed, while in administration, from the war zone.

3. Industry and Agriculture: The principle region of use of robots in farming is at the reaping stage. Organic product picking robots, driverless tractor/sprayer, and sheep shearing robots are intended to supplant human work. By and large, a great deal of elements must be considered (e.g., the size and shade of the organic product to be picked) before the initiation of an undertaking. Robots can be utilized for other green undertakings, for example, pruning, they ding, splashing and checking. Robots can likewise be utilized as a part of domesticated animal's applications (animals mechanical technology, for example, programmed draining, washing and maiming. Robots like these have numerous advantages for the farming business, including a higher nature of crisp produce, generation costs, and a littler requirement for physical work

VII. CONCLUSION

We here by conclude that our project can be useful in application such as transportation, inspection tasks industry and agricultural security and surveillance where in human intervention can be dangerous and can be avoided like mining, poisonous gas environment etc.future enhancement include additional of sensors like infra red,radar and in order to match the obstacles digital camera with high pixels can be used

REFERENCES

[1] Optimal Path Selection for Mobile Robot Navigation using genetic algorithm **Tamilselvi,Mercy shalinie,Hariharasudan** IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011

[2] Path Selection for Mobile Robots in Dynamic Environments**, Kristo Heero,Maarja Kruusmaa, Jan Willemson**

[3] Path planning of Autonomous Mobile robot **O.Hachour**, International journal of systems applications, engineering & development Issue 4, Volume 2, 2008

[4] Hybrid Approach for Global Path Selection & Dynamic Obstacle Avoidance for Mobile Robot Navigation **D. Tamilselvi, S. Mercy Shalinie, M. Hariharasudan and G. Kiruba**

[5] Grid-based Navigation for Mobile Robots,**Tucker** Navigation **D. Tamilselvi, S. Mercy Shalinie, M.**

PERFORMANCE ANALYSIS OF VARIOUS MANET ROUTING PROTOCOLS USING OPNET SIMULATOR

D.Chitra.[1] and Dr.M.Chandrasekaran[2]
1. Dept. of ECE, Mahendra Engineering College, 2. Dept. of ECE, Government College of Engineering, Bargur
drmcs123@yahoo.com, chitravalli2000@gmail.com

ABSTRACT- This paper provides a various mobile ad hoc network routing protocols namely AODV,DSR, OLSR and TORA and its performance in OPNET modeller. The mobile nodes were randomly placed in the network to give the possibility of multi hop routes from a node to the server. The protocols were simulated under varying conditions like node mobility, node density and with FTP traffic. The performance analysis of above mentioned routing protocols is determined with respect to routing overhead, throughput, end-to-end delay and packet delivery ratio. The performance of the routing protocol depends on various factors such as network load and mobility effects. In this paper, the performance of AODV, DSR, OLSR and TORA ad hoc routing protocols were evaluated in OPNET under varying network load and mobile speeds.

Keywords: MANET, AODV, DSR, OLSR, TORA, OPNET

I. INTRODUCTION

An ad hoc mobile network is a network which is a collection of randomly located wireless mobile nodes without the help of existing infrastructure network or centralized administration [1]. Due to the dynamic nature of nodes in the mobile ad hoc network, the network topology changes frequently which results in difficulty and complexity to routing among the mobile nodes within the network. These factors make the routing protocols vital in establishing communications among mobile nodes. In recent years, a lot of routing protocols and algorithms have been proposed and their performance under various traffic conditions and network environments studied and compared. The MANET provides a set of standard protocols which is robust and scalable to provide fast commercialization of mobile ad hoc networks in increasing network applications.

In general, ad hoc network routing protocols may be divided into two different categories. These are Proactive routing protocols and Reactive on-demand routing protocols[1]. Proactive routing protocols maintain up-to-date routing information between every pair of nodes in the network by sending, proactively, route updates at fixed time intervals. In proactive protocol, routing information is usually maintained in tables, the protocols are sometimes referred to as table-driven protocols. On the other hand, reactive on-demand routing protocols, , a route can be established to a destination only when there is a need, in general it is initiated by the source node through route discovery process within the network. Once a route has been established, it is maintained by the node until either the destination becomes inaccessible along every path from the source or has expired or until the route is no longer used.

Proactive protocols includes Destination-Sequenced Distance-Vector (DSDV) protocol, Cluster head Gateway Switch Routing (CGSR) protocol, Wireless Routing Protocol (WRP), Global State Routing (GSR), Optimized Link State Routing Protocol (OLSR), Fisheye State Routing (FSR) Protocol, Landmark Routing(LANMAR) Protocol, and Hierarchical State Routing (HSR).

Reactive routing protocols consist of Dynamic Source Routing (DSR), Ad-hoc on-Demand Distance Vector (AODV), Temporally Ordered Routing Algorithm (TORA), Associativity Based Routing (ABR), and Signal Stability Routing (SSR).

This paper is organized as follows: Section 1 dealt with the review the workings of the AODV, DSR, OLSR and TORA MANET routing protocols. Section2 provides the Performance metrics for evaluation. Section 3 gives simulation results and conclusions are given in Section 4.

II. Review of Routing Protocol Optimized Link State Routing (OLSR)

Optimized Link State Routing (OLSR) [2] protocol gives the stability of link state algorithm. In general, in a pure link state protocol, all the links with neighbour nodes are affirmed and are flooded in the entire network. But, OLSR is an optimized version of a pure link state protocol only designed for MANET. This protocol performs hop-by-hop routing that means, every node in the network uses its most recent information to route a packet. Therefore, even when a node is moving, the packets can be effectively delivered to it, if its speed is such that its movements could be followed in its neighbourhood. In this routing, the optimization is done mainly in two ways. Firstly, OLSR reduces the size of the control packets for a particular node by intimating only a subset of links with the node's neighbours who are its multipoint relay selectors, instead of all links network. Secondly, it reduces flooding of the control traffic by only using the selected nodes, called multipoint relays (MPR) to spread information in the network. Since only multipoint relays of a node can retransmit its broadcast messages, it considerably reduces the number of retransmissions in a flooding or broadcast procedure. Figure 1.1 shows a sample network structure used in OLSR.OLSR protocol relies on the selection of multipoint relay nodes. Each node calculates the routes to all known destinations through these nodes. These MPRs are selected among the one hop neighbourhood of a node

using the bidirectional links, and they are used to reduce the amount of broadcast traffic in the network.

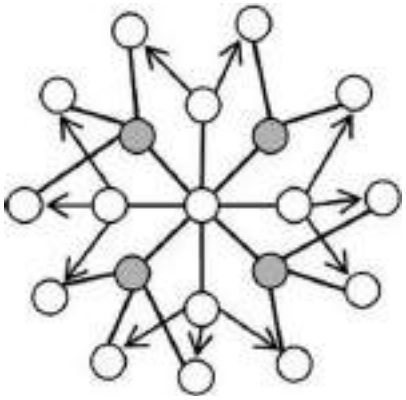
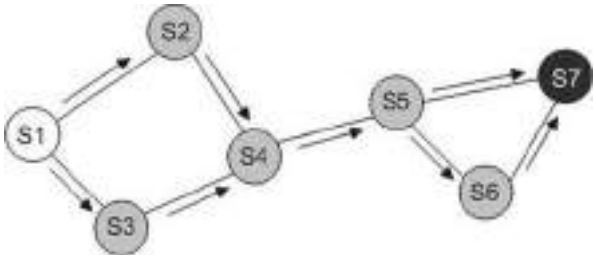


Fig. 1.1 Multipoint Relays (MPRs) are in gray colour. The transmitting node is shown at the center of the sample structure

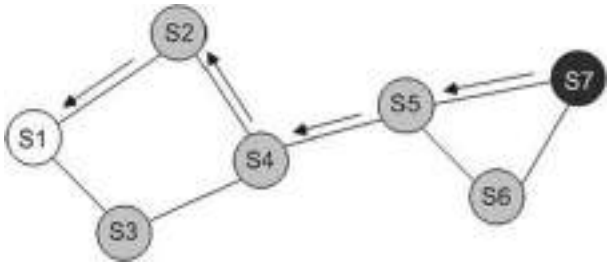
Ad Hoc On-Demand Distance Vector Routing(AODV)

Ad Hoc On-Demand Distance Vector Routing [3] is a reactive type routing protocol. It reduces the number of broadcasts by creating routes based on need. When any source node wants to send a packet to a destination, it starts the process and it broadcasts a route request (RREQ) packet. The neighbouring nodes sequentially broadcast the packet to their neighbours and the process continues until the packet reaches the destination. During the forwarding process of the route request, intermediate nodes trace the address of the neighbour from which is the first copy of the broadcast packet is received. This record is saved in their route tables, which helps for establishing a reverse path. If extra copies of the same RREQ are received later, these packets are discarded. By using the reverse path the reply message is sent. For route maintenance phase, when a source node leaves, it re-initiates a route discovery procedure. If any intermediate node moves within a particular route, the neighbour of the drifted node can identify the link failure and gives a link failure announcement to its upstream neighbour. This procedure prolonged until the failure notification attains the source node. Based on the received information, the source may decide to re-initiate the route discovery phase.

Figure 1.2 shows an example of AODV protocol operation details. In this Figure, S1 is the source node and S7 is the destination node. The source node initiates the route request and the route is created based on need. Route reply is sent using the reverse path from the destination node.



(a)

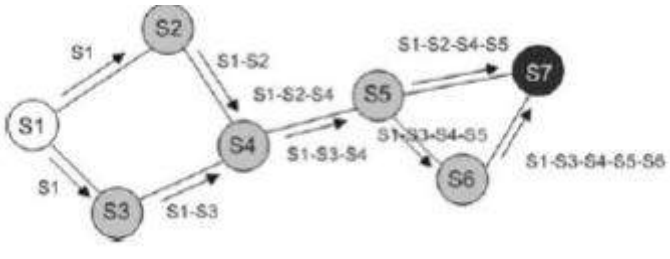


(b)

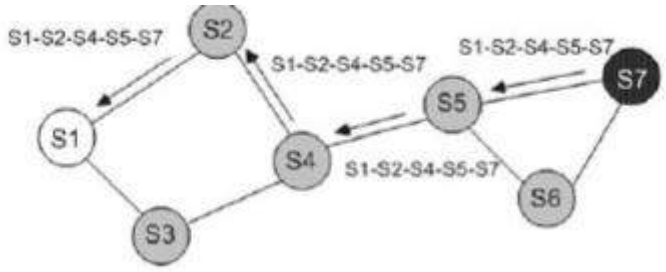
Figure 1.2 AODV protocol (a) Source node broadcasting the route request packet. (b) Route reply is sent by the destination using the reverse path

Dynamic Source Routing (DSR)

Dynamic Source Routing [4] allows nodes in the MANET to dynamically cover a source route across multiple network hops to any destination. In DSR protocol, the mobile nodes are required to preserve route caches or the known routes. Update the route cache, when any new route is known for a particular entry in the route cache. Routing in DSR is done by using two phases. One route discovery phase another route maintenance phase. When a source node desires to send
a message packet to a destination, it first consults its route cache to determine whether it already knows about any route to the destination or not. If already there is a route entry for that destination, the source node uses that to send the packet. Otherwise, it initiates a route request broadcast process. This route request includes the source address, destination address and a unique identification number. Every intermediate node verifies whether it knows about the destination or not. If the intermediate node does not know about the destination, it forwards the packet another time and finally this reaches the destination. A node processes the route request packet as long as it has not previously processed the packet and its address is absent in the route record of the packet. Destination or any of the intermediate nodes generate the route reply when it knows about how to reach the destination. Figure 1.3 shows the operational method of the dynamic source routing protocol. In Figure, the route discovery procedure is shown where S1 is the source node and S7 is the destination node.



(a)



(b)

Fig. 1.3 (a) Route Discovery (b) Using route record to send the route reply

In this example, the destination gets the request through two ways. It chooses one path based on the route records in the incoming request packet and accordingly sends a route reply using the reverse path to the source node. At each hop, the best route with minimum hop is stored. In this example, the route record status at each hop showed to reach the destination from the source node. Here, finally the chosen route is S1-S2-S4-S5-S7.

Temporarily Ordered Routing Algorithm (TORA)

Temporally Ordered Routing Algorithm (TORA) [5] is also a another type of a reactive routing protocol which is some improvements in proactive type in which a link between nodes is formed by creating a Directed Acyclic Graph (DAG) of the route to the destination node from the source node. In route discovery phase, it applies a link reversal model. A route discovery query is broadcasted and throughout the entire network it is propagated until it ends the destination or a node that has information regarding how to reach the destination. TORA provides a parameter, named as height. Height is a distance of the responding node's distance up to the mandatory destination node. In route discovery process, this parameter is return back to the querying node. As the query reply propagates backside, each intermediate node updates its information in the TORA table with the route and its height to the destination. Then, the source node selects the best route by using height parameter toward the destination. This protocol has an important property is that it frequently chooses the most appropriate route, rather than the shortest route. For all these, TORA tries to minimize the routing traffic overhead.

III. PERFORMANCE METRICS FOR EVALUATION

In this paper, four performance metrics are used to determine the overall network performance. These metrics are namely routing overhead, packet delivery ratio, packet end-to-end delay and network throughput [9].

1. Routing Overhead

As the network grows, various routing protocols perform differently. The amount of routing traffic increases as the network grows. This parameter measures the scalability of the protocol, and thus the network. It is defined as the total number of routing control packets transmitted over the network, which is expressed in bits per second or packets per second. Some sources of routing overhead in a network are cited in [6] as the number of neighbours to the node and the number of hops from the source to the destination. Other causes of routing overhead are network congestion and route error packets.

2 Packet Delivery Ratio

Packet Delivery Ratio (PDR) is the ratio between the number of packets generated at the sources to the number of packets received by the destination. This metric reflects the network throughput. Thus, this metric is useful to measure any degradation in network throughput. A high packet delivery ratio is desired in a network.

3. Packet End-to-End Delay

The packet end-to-end delay is the time from the generation of the packet by the sender up to their reception at the destination. It is expressed in seconds. This metric includes not only the delays of data propagation and transfer, but also all possible delays caused by buffering, queuing and retransmitting data packets[10].It represents the reliability the routing protocol.

4.Throughput

It is defined as the ratio of the total amount of data that reaches a receiver from a sender to the time it takes for the receiver to get the last packet is referred to as throughput [7]. It is expressed in bits per second or packets per second. Frequent topology changes, unreliable communication, limited bandwidth and limited energy these factors that affect the throughput in MANETs [7]. A high value of throughput network is desirable.

SIMULATION RESULTS

In this paper, the AODV,DSR, OLSR and TORA were analysed with the help of Opnet Modeller. This paper aims at modelling the behaviour of the routing protocols under varying network loads and speeds. Global discrete event statistics (DES) on each protocol and wireless LAN[8] were collected. Therefore average statistics of the throughput, delay, packet delivery ratio and routing overhead for various node size like 10, 20, 50 mobile nodes for the entire MANET were examined using Opnet Simulator. Main characteristics of the scenarios maintained are depicted in the Table 1.1.

Table 1.1 : Simulation Model

Statistics	Value
Scenario Size	1 X 1 Km
Simulation Time	1 h
Nodes	10,20,50
802.11 data rate	11 Mbps
Mobility Model	Random Waypoint
Application type	FTP
Transmission Power	0.005W
Pause time	300s

Routing overhead, packet end-to-end delay and the throughput of the network were analysed. Global statistics

for the entire network were collected and average values were given. Global statistics for TORA was unable to collect with higher traffic sources, i.e. 50 nodes. Because of TORA protocol took more computer memory usage during the simulations. AODV,DSR,TORA and OLSR protocols were simulated with various scenarios is a combination of 10, 20 and 50 mobile nodes moving at constant speeds of 10 m/s and 30 m/s.

Figure 3.1 : Opnet simulation setup

Routing Overhead

.

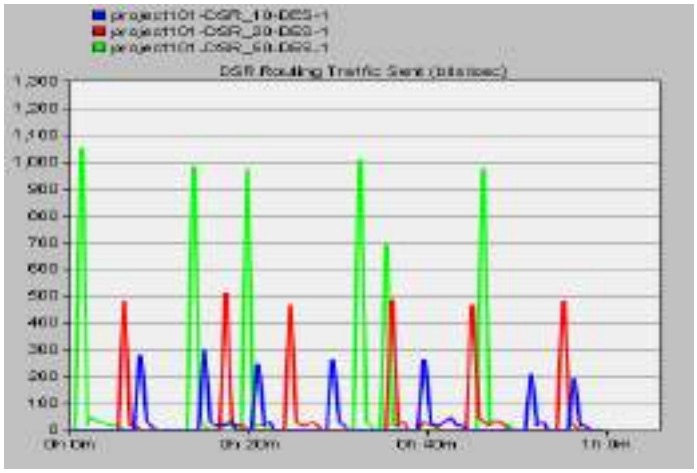


Figure 3.2 : Routing overhead in AODV and DSR for 10,20 and 50 nodes

In both DSR and AODV high mobility implies that there are frequent link breakages.

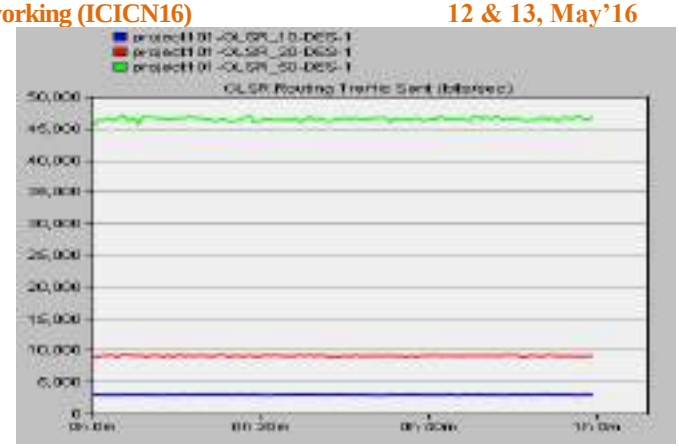


Figure 3.3: Routing overhead in OLSR and TORA for 10, 20 and 50 nodes

In OLSR, the increasing the mobility has no effects on the amount of routing traffic injected into the network. In TORA networks with large traffic sources, it performs better at higher than at lower mobility.

It was observed that OLSR sends the highest amount of routing traffic into the network followed by TORA, AODV and DSR. In DSR with the least amount of routing traffic sent. So, in routing overhead, DSR outperforms AODV, TORA and OLSR as it sends the least amount of routing traffic into the network

Packet End-to-End Delay

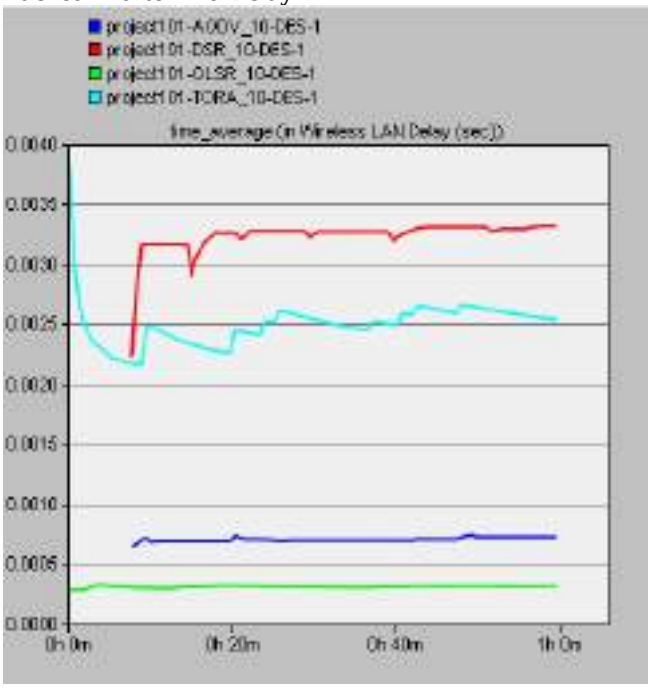


Figure 3.4 : End-to-end delay –10 sources at 10m/s , 20 sources at 30m/s

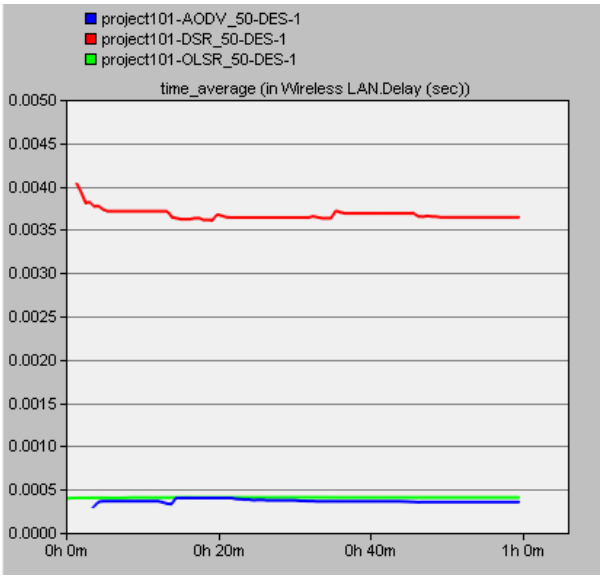


Figure 3.5 : End-to-end delay –50 sources at 30m/s

It was observed that OLSR having very low delay under light and medium load conditions. When heavy load condition AODV had low delay and mobility did not have an effect on the delay. DSR had a higher end-to-end delay as the network density is larger.

Packet Delivery Ratio

TORA delivered more number of packets with low speed and low number of traffic sources. When the number of sources increased, it was degraded from highest to lowest. TORA had the least packet delivery ratio when the nodes had a speed of 30 m/s with low number of traffic source. This increased as the number of nodes increased to 20.

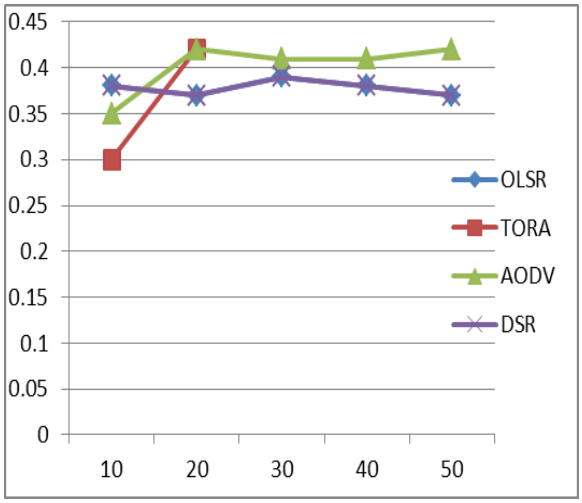


Figure 3.6: Packet delivery ratios at 10 m/s

At low speeds, AODV outperformed both DSR and OLSR in the networks with 5 and 20 traffic sources. When the traffic sources increased to 50, the packet delivery ratio for AODV degraded significantly and was comparable to that of DSR. OLSR at this stage outperformed all the other protocols. AODV outperformed OLSR and DSR in the larger network when the nodes were moving at 30 m/s. In the smaller network i,e less number of nodes, DSR and OLSR outperformed comparing AODV.

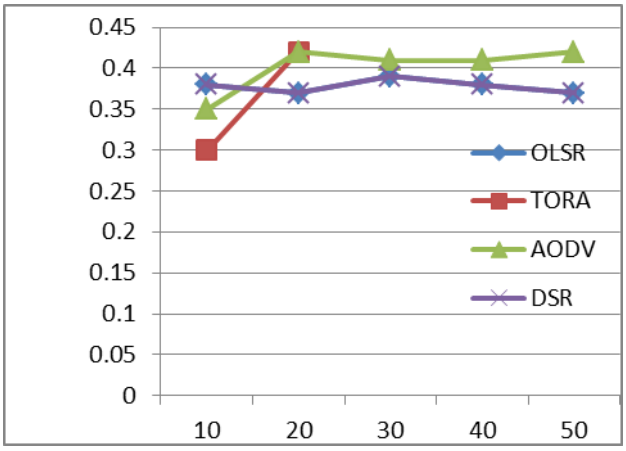


Figure 3.7: Packet delivery ratios at 30 m/s

Throughput

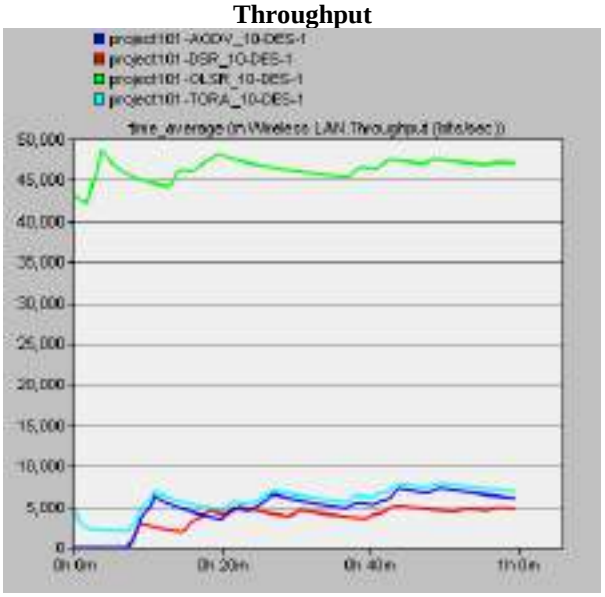


Figure 3.8: Throughput –10 sources at 10m/s , 20 sources at 30m/s

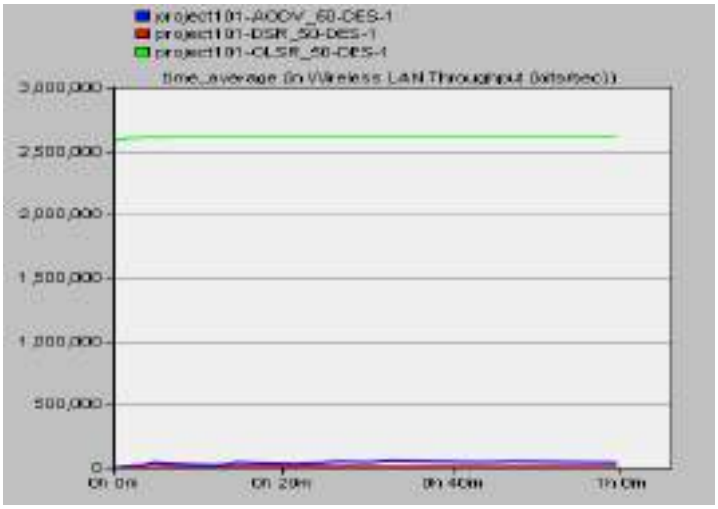


Figure 3.9: Throughput –50 sources at 30m/s

From the results, DSR and AODV provides better throughput than OLSR in smaller as well as larger

networks both in low and high-speed scenarios. The throughput performance of TORA at large network load cannot be obtained because no data available.

IV. CONCLUSIONS

In this paper, four different ad hoc routing protocols AODV, DSR, OLSR and TORA deployed over MANET were examined using FTP traffic analyzing the behaviour with respect to their routing overhead, throughput and packet end-to-end delay. From the above analysis, it is concluded that AODV had low delay when heavy load condition. Throughput of DSR and AODV is better as compared to OLSR. In routing overhead DSR performs well than AODV, TORA and OLSR.

From this analysis, it is concluded that among the considered protocols, there is no single one with an overall superior performance. One protocol was superior in terms of routing overhead whilst others were superior in terms of packet end-to-end delay, packet delivery ratio, or throughput. The choice of a particular routing protocol will depend on the intended use of the network.

References

[1] Stefano Basagni, Marco Conti, Silvia Giordano, Ivan Stojmenovic, "Mobile ad hoc networking", John Wiley & Sons, inc., publication,2004.

[2] Jacquet P, Mu¨ hlethaler P, Clausen T, Laouiti A, Qayyum A, Viennot L (2001) Optimized Link State Routing Protocol for Ad Hoc Networks. IEEE INMIC 2001:62–68

[3]. Perkins CE, Royer EM, Chakeres ID (2003) Ad hoc On-Demand Distance Vector (AODV) Routing. IETF Draft, October, 2003.

[4]. Broch J, Johnson DB, Maltz DA (1999) The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks. IETF Draft, October, 1999.

[5].Park VD, Corson MS (1997) A highly adaptive distributed routing algorithm for mobile wireless networks. Proceedings of IEEE INFOCOM 1997, Volume 3:1405–1413.

[6].Mohammad Naserian, Kemal E. Tepe and Mohammed Tarique, "Routing overhead analysis for reactive routing protocols in wireless ad hoc networks," IEEE Conference on Wireless And Mobile Computing, Networking And Communications, WiMob, 2005, pp. 87 – 92.

[7].Uyen Trang Nguyen and Xing Xiong, "Rate-adaptive Multicast in Mobile Ad hoc networks," IEEE International Conference on Ad hoc and Mobile Computing, Networking and Communications, WiMob, Montreal, Canada, 2005.

[8].Bained Nyirenda ,Jason Mwanza," Performance Evaluation of Routing Protocols in Mobile Ad hoc Networks", 2009.

[9].Diya Naresh Vadhwani, Deepak Kulhare, Megha Singh, Behaviour analysis of DSR MANET protocol with http traffic using Opnet simulator , International Journal of Innovative Research in Computer and Communication Engineering,Vol.1,Issue 3, May 2013.

[10].Sumedha S. Deshpande, Krutika V. Asare, Swapnil Deshpande, "An overview of Mobile Ad Hoc Networks for the Proactive, Reactive and Hybrid Routing Protocol", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 4 April, 2013 Page No. 1234-1243

Offline Handwriting Identification Using Adaptive Neural Fuzzy Inference System

Devi T, Swathi Priya N, Prabaharan J, Manigandan J
Department of Computer Science and Engineering, Rajarajeswari College of Engineering, Bangalore-74
deviyabe@gmail.com, swathinatarajan@gmail.com, prabaharan.jothi@gmail.com, jmaniannauniv@gmail.com

ABSTRACT - A new method for handwriting identification was presented. The chosen word from the text image (handwritten) which is given as input to the system is separated as individual characters. Then each of the separated characters are converted into 625 values from column vectors, and then fed into the adaptive neural fuzzy inference system (ANFIS) for membership functions (MF) calculation and normalized firing strength. In our paper we were used triangular membership function and compare with others MF.The networks has been designed with single layered neural network corresponding to a character from a-z, the outputs of all the column vector is fed into network the which has been developed using the concepts of correlation. The output is compared with genetic algorithm and results are approximate to the scale, so ANFIS network is efficiency.

Keywords—*Handwriting Identification, Adaptive Neuro Fuzzy Inference System (ANFIS), Column vector*

I INTRODUCTION

The identification of written characters can be done using Handwriting recognition. The problem can be viewed in such a way to identify the most appropriate character the given figure matches to. Recognition technique is used for offline character recognition refers to the where the final figure is given to us. Online character recognition systems are contrary to the where the data can be sampled while the character is being written. Operating in offline mode gives input in the complete picture character that we need to recognize. The recognition is usually in complexity which is related with the size of the language being considered. If the language contains more number of characters; difficulty arises in identification than the case when the language contains lesser number of characters. They always have an effect on handwriting recognition system. In this paper we propose the use of Column vectors for solving this problem. The basic idea of column vector comes from the fact that it can be used as an excellent means of combining various styles of writing a character and make way to new styles. Closely observing the capability of human mind in the recognition of handwriting, we find that humans are able to recognize characters even though they might be seeing that style for the first time. This is possible because of their power to visualize parts of the known styles into the unknown character. In this paper we try to depict the same power into the machines.

The problem increases when it operates offline mode. We see a lot of work has been done in this area in the past few years. The solutions being proposed mainly use Adaptive Neural Fuzzy Inference System (ANFIS) and Hidden Markov Models (HMM) for solving the problem. Column vectors have not been applied much. They have been applied for feature selection optimization. ANFIS Neural system involve training of the system with all the characters [3].Then when an unknown input is given to the system, the ANFIS Neural Network is able search and find probable character by generalization. Hence once trained, the system would be ready to recognize input which is unknown. Hidden Markov Model is a complete statistical model that tries to predict the unknown sequence. It also tries to recognize the unknown character which is given as input.

II COLUMN VECTOR

Column vectors are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to these structures as to preserve critical information.

Column vectors are a very good means of optimizations in such problems. They optimize the desired property by generating approximate solutions from the presently existing solutions. Column vectors are optimization heuristics. It is used to search for good solution to complex problem and good heuristics for combinatorial problems. Mutation and crossover may only applicable for bit string and integer representation. It generates hybrid solution from the existing solution. This solution may be better than the already existing solution.

A. General Procedure

Handwriting recognition is a famous problem like the recognition of whatever input which is given in the form of image, scanned paper, etc. The handwriting recognition generally involves the following steps [5].

B. Segmentation

This step deals with the breaking of the lines, words and finally getting all the separated characters and the identification of the boundaries of the character, separating them for further processing. We assume that this step is already done using this algorithm. Hence the input to our system is a single character.

C. Preprocessing

It involves the initial processing of the image, so that it can be used as an input for the recognition system. In this algorithm a part of preprocessing step has been done and (thin) character segment is made to a unit pixel thickness. Various algorithms may be used for this purpose. The further processing is done by our algorithm.

D. Recognition

Recognition which identifies the character, once the input image is available in good condition. It may be processed for recognition which identifies the character. This algorithm uses an image as an input for the same.

i= imread('d:/filename.bmp')

To store the input image in bitmap file format that input image then again stored into variable i. Then calculate the column and row values for corresponding image. These vector values converted into the single row, similarly calculate covariance of same single row vector values. Then calculate intensity value for that corresponding image. These two inputs are given to ANFIS neural networks.

i(:)

s=cov (i(:))

intensityValue = i(row,column)

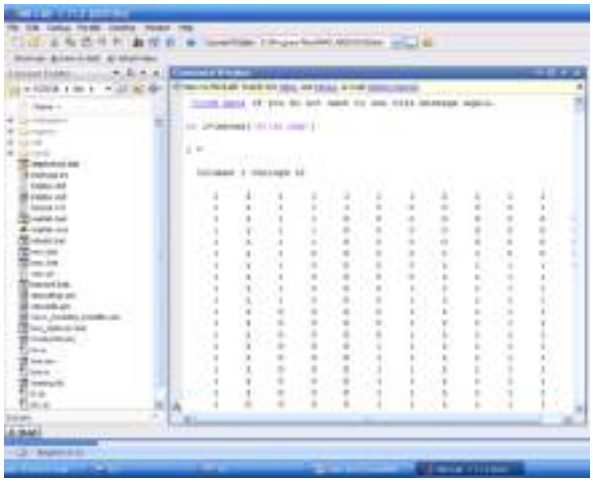


Fig. 1: Column Vector

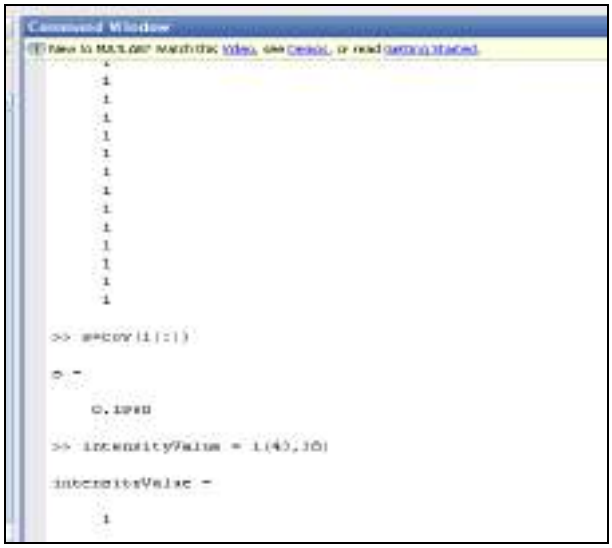


Fig. 2: Intensity Value

III ANFIS

ANFIS stands for Adaptive Neuro Fuzzy Inference System. ANFIS are a class of adaptive networks that are functionally equivalent to fuzzy inference systems. It integrates both neural network and fuzzy logic principles. ANFIS is a universal approximator. It's trained by a hybrid learning algorithm. Using a given input/output data set, the toolbox function ANFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned.

The ANFIS algorithm combines the best features of both artificial neural network (ANN) and fuzzy logic.

The input data is trained here and the ANFIS toolbox is a five layered network. There are rules through which the nodes combine and give the output in a better way and compare with genetic algorithm. Fig1 shows the various layers of ANFIS and the combination of nodes.

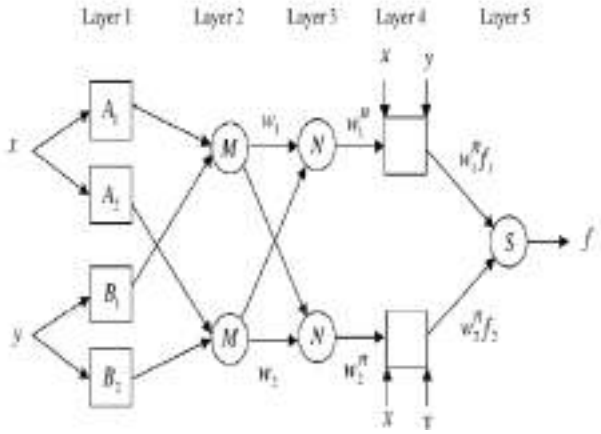


Fig. 3: ANFIS Architecture

In layer 2, we calculate membership function. Determine the function of F by using triangular fuzzy membership. The general expression of triangular fuzzy member function is,

$$f(x,a,b,c)=\max\left(\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),0\right) \quad \text{(equation 1)}$$

Where a,b,c are parameter set.

IV ANFISGA

This method brings the technique of ANFIS and GA together. The dataset which is obtained by the segment characters for training and testing. A set of data is trained by ANFIS and this is done in ANFIS Graphical User Interface of MATLAB. Here we use hybrid optimization method and error tolerance is set to zero. The membership function used is triangular membership function. The data to be evaluated is fed into the optimization toolbox. The various operations of the column vector are performed and the fitness function is evaluated until the stopping criterion is met. Fitness of each chromosome is computed by minimum optimization method. The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, and is defined by the formula:

$$M=\frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right| \quad \text{(equation2)}$$

Where A_t is the actual value and F_t is the forecast value.

The difference between A_t and F_t is divided by the actual value A_t again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. multiplying by 100 makes it a percentage error.

Fig. 3 shows the operation of ANFIS and the Column vector operations and the output is the dosage level which is produced as per the given data.

The membership function used is triangular membership function. It is calculated as in the following equation1.

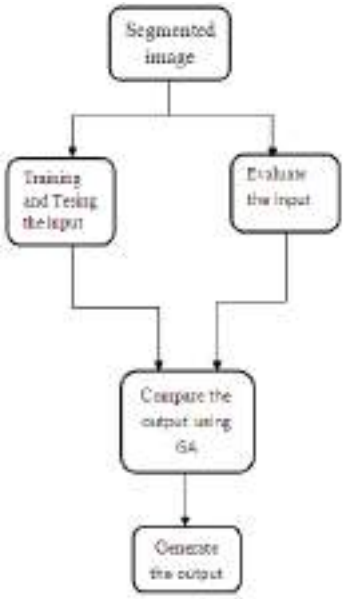


Fig. 4: ANFISGA data flow diagram

V RESULTS

The training and testing is done in ANFIS toolbox of MATLAB. The results show that this technique is much better than other statistical techniques and artificial intelligence techniques such as artificial neural network and fuzzy logic where the accuracy is low and a lot of risk is associated.

This model is implemented till the ANFIS and an accuracy of 93% is achieved which is much accurate when compared to the previous models. These results are stored and then evaluated by Column vector.

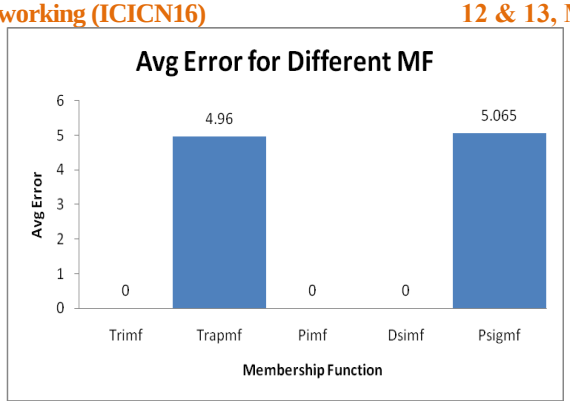


Fig. 5: Average error for different MF

VI CONCLUSION

In our project improve efficiency for handwriting recognition has been proposed. The proposed system is expected to produce better accuracy than the existing system. The result is efficient while comparing with genetic algorithm. We proposed the use of column vector for solving the problem of offline handwriting recognition. Input is taken in the form of images. Training data is taken to train the algorithm and it was already present in the database.

REFERENCES

[1] Rahul KALA1, Harsh VAZIRANI2, Anupam SHUKLA3 and Ritu TIWARI,' Offline Handwriting Recognition using Genetic Algorithm' IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010

[2] FLink, Gernot A and Plotz, Thomas, 'Unsupervised Estimation of Writing Style Models for Improved Unconstrained Off-line Handwriting Recognition', International Workshop on Frontiers in Handwriting Recognition, Suvisoft, France, 2006

[3] Tripathy N. and Pal U,'Handwriting segmentation of constrained Oriya text', Sadhna,2006.

[4] Som, Tanmoy and Saha, Sumit, 'Handwritten character recognition by using Neural-network and Euclidean distance metric', Social Science Research Network, 2008.

[5] Liwicki, Marcus and Bunke, Horst, 'Int. Journal of Pattern Recognition and Artificial Intelligence, 2007, 83-98.

[6] Graves, Alex; Fernandez, Santiago; Liwicki, Marcus; Bunke,Horst and Schmidhuber, Jurgen, 'Unconstrained Online Handwriting Recognition with Recurrent Neural Networks',Advances in Neural Information Processing Systems 20,2008.

DETECTION AND IDENTIFICATION OF LBBB AND RBBB RHYTHMS IN ECG WAVES USING GABOR TRANSFORM ANALYSIS

Geetha A, T.R Gopalakrishnan Nair, Asharani M
[2]RECTOR, RR Group of Institutions, Bangalore, India.

ABSTRACT -Automated analysis of ECG signals has evolved over a period of time to provide very sophisticated system analysis and interpretations. As the demand in the medical field is moving towards fully automated and intelligent systems, better technology and methods are required to interpret the abnormalities of body, especially of heart. Abnormal rhythms like LBBB and RBBB may not be life threatening but an early detection of these can avoid further complications. Here we present a signal processing approach to detect and identify L, R rhythms in ECG signals for early warning. Gabor transform technique is used for the analysis of the signal and the approach is capable of detecting the abnormality with sufficient accuracy.

Index Terms— ECG, LBBB, RBBB, STFT, Gabor Transform

1. IINTRODUCTION

Electrocardiography (ECG) is the process of recording the electrical activity of the heart using sensitive electrodes. An ECG signal can provide a good insight into the functioning of the heart. The effective depolarization of the ventricle mass cells requires that there be a specialized conduction pathway within the ventricle. Ventricular conduction system is divided into right and left Bundle branches after Bundle of His. Blocks can occur in this bundle branches and can results in Left bundle branch block (LBBB) and Right bundle branch block (RBBB). In RBBB the right ventricle is activated with a time delay after left ventricle. Because of this, the contraction of ventricle is also delayed and this can be observed as a significant change in ECG. QRS signal will be prolonged and there will be appearance of 'bunny ears' in V1 and V2 leads. In LBBB the left ventricle depolarizes with small time delay after the right ventricle. The QRS will be prolonged with Leads I and V6 will record long R waves. A larger R' like an odd pair of 'Bunny ears will be visible. In v1 lead, the R wave will be very small and S wave will be prolonged. It may become 'Bunny eared'. In both the cases T- wave will be abnormal. The block can be caused by coronary artery disease, cardiomyopathy, or valve disease. Even though the bundle branch block itself may not be a problem, it can be a warning sign of other, more serious heart conditions. How to detect the occurrence of RBBB and LBBB in an automated way remains to
be challenge and very deep investigations are on way. The detection of phenomenon based digital signal processing techniques are under development in various laboratories. In order to solve this Gabor Transforms can be a candidate technology. The Gabor transform is a special case of the short-time Fourier transform (STFT). This transform can be used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. In STFT the signal to be analyzed is divided into frames and each frame is Fourier transformed, which provides magnitude and phase for each segment. As the window is moved along time we will get a two dimensional matrix in time – frequency domain. $STFT\{x(n)\}(m,\omega) = \sum_{n=-\infty}^{+\infty} x[n]w[n-m]e^{-j\omega n}$.
The magnitude squared of the STFT yields the spectrogram of the function. In Gabor transform the window used is a Gaussian function and the resulting function after multiplication is then subjected to Fourier transform to derive the time-frequency analysis. The window function means that the signal near the time being analyzed will have higher weight. The Gabor transform of the signal x(t) is given by $G_x(t,f) = \int_{-\infty}^{\infty} e^{-\pi(\tau-t)^2} e^{-j2\pi f\tau} x(\tau)d\tau$. The Gaussian window parameters can be adjusted to extract the required features.

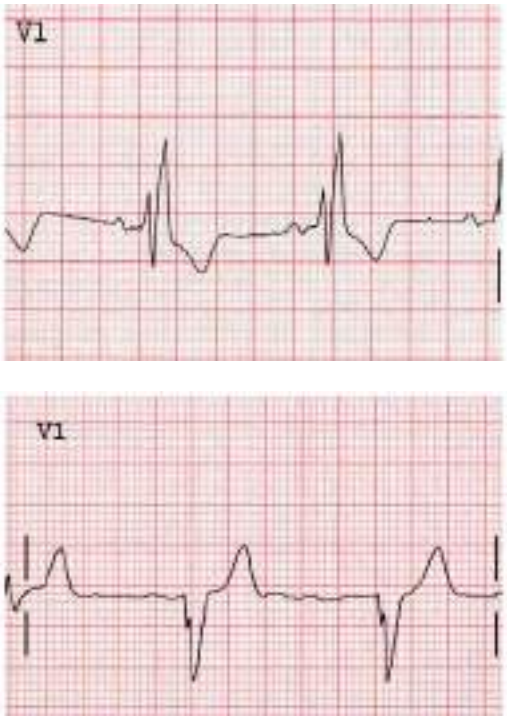


Fig1. (a) RBBB (b) LBBB

Yu et. Al [1] has used discrete wavelet transform for feature extraction. Three sets of statistical features of the decomposed signals as well as the AC power and the instantaneous RR interval of the original signal are exploited to characterize the ECG signals. In [2] authors have used

QRS pattern recognition method for computation of a large collection of morphological QRS descriptors and Matching Pursuits algorithm for classification algorithm, Huang et.al has developed a method to classify LBBB and RBBB from normal beats. They used a combination of three different types of classifiers: a minimum distance classifier constructed between NORM and LBBB; a weighted linear discriminant classifier between NORM and RBBB based on Bayesian decision making using posterior probabilities; and a linear support vector machine (SVM) between LBBB and RBBB. Their method predicted a sensitivity of 91.4% for LBBB and 92.8% for RBBB. In [2] authors have extracted the features like QRS duration, R-R interval etc. and used a Multilayer Perceptron (MLP) classifier. Amandeep et.al used stationary wavelet transform to extract the features for classification.

2. RESEARCH BACKGROUND

Automatic analysis of ECG is gaining momentum over a period of time. With the advancement in technology faster and accurate methods of analysis is required for meeting the present day requirement in telemedicine and automated Biosignal analysis. Even though LBBB and RBBB are not a life threatening abnormality, it is an indicator of hidden problems related to heart. A timely detection can prevent further complications. There are many methods which were adopted for the classification of beats of normal rhythm, LBBB, RBBB, PVC etc. Most of these methods adapted Fourier transform or wavelet features for classification. In most of these cases neural network, principal component analysis, and different classification algorithms are used. But in case of LBBB and RBBB, the major requirement is to know the presence of LBBB and RBBB over normal rhythm which is sufficient for further investigation. After detecting LBBB or RBBB, deeper analysis may be required to find the hidden cause, by seeking the expert advice of a physician.

In this work, our aim was to detect the presence of LBBB and RBBB in comparison to normal rhythm. For this we have used the time frequency analysis STFT using Gaussian window. Gaussian window parameter center frequency and standard deviation give the flexibility to extract the required parameters from ECG waves.

3. RESEARCH METHODOLOGY

The required data for analysis is collected from the ECG signals arranged for analysis. It is collected from global data source where annotated ECG signals are available. MIT-BIH Arrhythmia diagnostic database was used for the testing of algorithm. MIT-BIH Arrhythmia Database provides a collection of 48 fully annotated half-hour two-lead ECGs. It also provides a database directory with a table of beat types. This provides enough data for testing and since each beat is annotated the verification was easy. ECG data strips for 10 sec were collected and each one was analyzed separately. Before any ECG feature extraction the first step should be to remove the noises. The main noise which can affect the detection is the baseline wandering. So we have used median filters to remove baseline drift.

Baseline Drift Removal

Any ECG signal analysis requires a pre-processing to remove the noise artifacts. For a computerized detection of ECG features the main affecting noise factor will be the baseline drift. The frequency content of the baseline wander is usually in a range well below 0.5Hz. This baseline drift can be eliminated using median filters (200-ms and 600-ms). The original ECG signal was processed with a median filter of 200-ms width to filter QRS complexes and P waves. The resulting signal was then processed with a median filter of 600-ms width to remove T waves. The signal resulting from the second filter operation contained the baseline of the ECG signal, which was then subtracted from the original signal to produce the baseline corrected ECG signal. The baseline corrected signal for the ECG signal 108 V1 lead is shown in fig 2. The intermediate results of unravelling QRS & P-wave as well as T-waves are also shown in Fig 2.

The base line corrected signal is then subjected to short time Fourier Transform using Gaussian window. Simulation was carried out in MATLAB. Gaussian window with varying standard deviation and centre frequency were used and found the optimum values to get a spectrogram which enhances the LBBB and RBBB frequency bands.

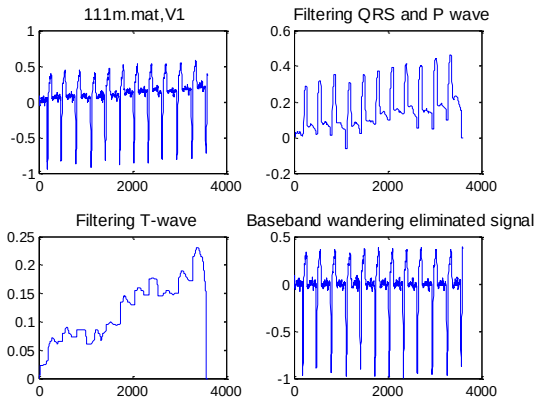
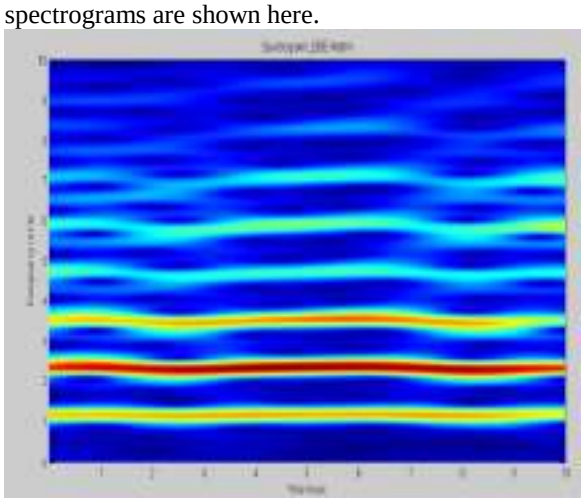


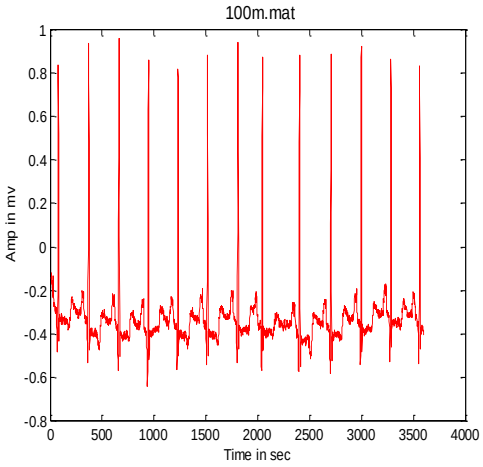
Fig 2. Base drift eliminated signal

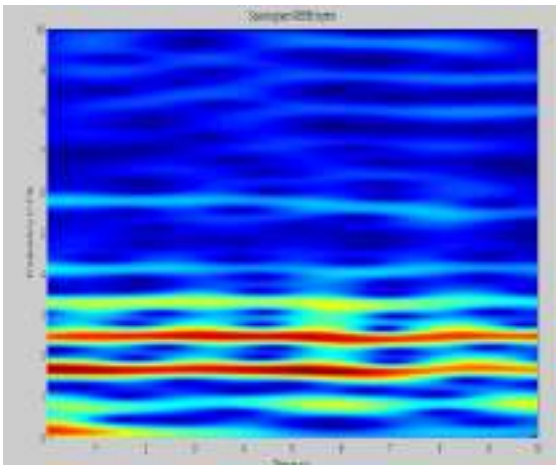
4. RESULT

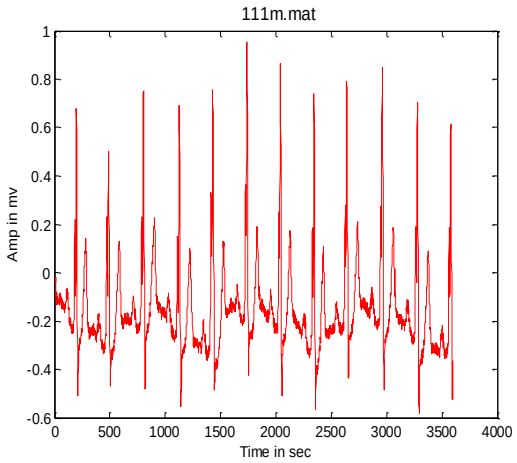
The proposed algorithm was tested on MIT-BIH arrhythmia database. MIT-BIH database consists of 9 records, each of 30 minutes duration with two simultaneous channels sampled at a frequency of 360Hz which has LBBB and RBBB beats. Each record was tested for LBBB and RBBB presence and found to be detecting it. Spectrogram was able to give the location and the presence of each LBBB and RBBB over normal rhythm. Fig.3. Below gives the spectrogram of three signals with LBBB, RBBB and normal rhythm. From the spectrogram it can be seen that for RBBB the significant frequencies are less than three Hz. In LBBB case the significant spectrogram frequency is above 4Hz. In normal rhythm case the significant frequencies are above 10Hz. Here the Gabor Transform analysis over ECG signal yielded a comprehensive frequency perspective of the performance of heart indicating vital deviation from normal performance in a graphical way. Fig 3 shows the three different cases considered for analysis. A normal waveform,

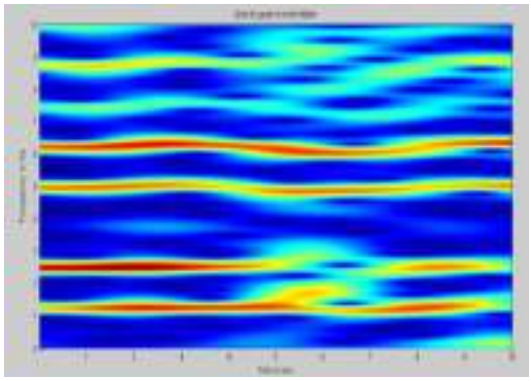
LBBB and RBBB rhythms with their corresponding spectrograms are shown here.











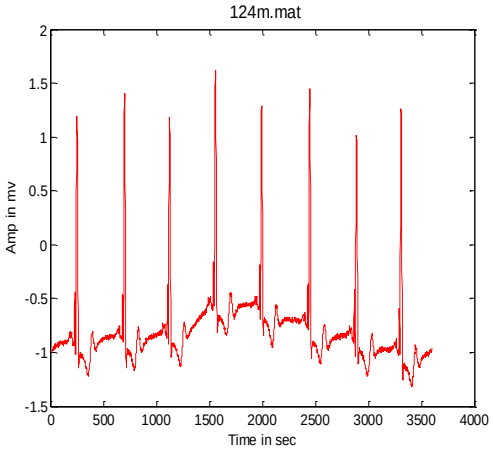


Fig 3. a) LBBB b) Normal Rhythm c) RBBB

5. CONCLUSION

The method of applying Gabor Transform for ECG signal analysis has enabled the detection of time sequenced frequency fluctuations occurring in ECG signals. The method of identifying LBBB, RBBB rhythms as presented here was found to be efficient for detecting those phenomena. This yielded a definite improvement on ECG analysis bringing viability to automated detection of 'L R' rhythms. More study is required to identify the further scope of this approach by closely observing ECG variations.

REFERENCES

[1] Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Sung-Nien Yu, ,Ying-Hsiang Chen, Pattern Recognition Letters, Volume 28, Issue 10, 15 July 2007, Pages 1142–1150

[2] Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification, Ivaylo Christov , Gèrman Gómez-Herrero, Vessela Krasteva, Irena Jekova, Atanas Gotchev, Karen Egiazarian, Medical Engineering & Physics Volume 28, Issue 9, November 2006, Pages 876–887

[3] Discrete Gabor Transform S Qian, D Chen - Signal Processing, IEEE Transactions on, 1993

[4] Detection and classification of P waves using Gabor wavelets, M Michaelis, S Perz, C Bla, G Sommers International congress Cardiology, London 1993 ,531-54

[5] Gabor Feature Extraction for Electrocardiogram Signals, Gwo Giun (Chris) Lee, Jhen-Yue Hu, Chun-Fu Chen, Huan-Hsiang Lin

[6] Asymptotic Wavelet and Gabor Analysis: Extraction of instantaneous frequencies. N. Delprat , B. Escudi´e P. Guillemain , R. Kronland-Martinet , Ph. Tchamitchian , B. Torr´esani, IEEE Transactions on Information Theory · April 1992,

[7] Detection of inter-patient left and right bundle branch block heartbeats in ECG using ensemble classifiers. H Huang, J Liu, Q Zhu, R Wang, G Hu - 2014 - biomedcentral.com

[8] Classification of Right Bundle Branch Block and Left Bundle Branch Block cardiac arrhythmias based on ECG analysis, Sukanta Bhattacharyya, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 3, March 2014.

[9] LAFB, LBBB, RBBB and Myocardial Infarction Detection Using 12 Lead ECG, Amandeep Kaur , Mohit Arora, IJRECE VOL. 3 ISSUE 3 JULY-SEPT. 2015 ISSN: 2393-9028 (PRINT) | ISSN: 2348-228

[10] Total Removal of Baseline Drift from ECG Signa,V.S. Chouhan and S.S. Mehta, Proceedings of International conference on Computing: Theory and Applications, ICTTA-07, pp. 512-515, ISI, Kolkata, India, 2007.

[11] ECG baseline wander correction by mean-median filter and discrete wavelet transform, W. Hao, Y. Chen, and Y. Xin, IEEE Engineering in Medicine and Biology Society, 2011

[12]www.physionet.org/physiobank

Review Paper on Optical Random Access Memory

Jyoti.B
Assistant Professor, Dept of Electronics and Communication Enginering, GSSSIETW, Mysuru
jyothi.b@gsss.edu.in

ABSTRACT-Optical communication has the potential to fulfill the nowadays bandwidth requirements, provided that it can be implemented in a cheap and reliable way. Development of optical communication is hindered by lack of all Optical memory which is essential for storing information, retiming, regeneration, and reshaping. A optical memory can greatly enhance both capacity and flexibility of optical network. Optical Random-Access Memory (o-RAM) has been regarded as one of the most difficult challenges in terms of replacing its various functionalities in electronic circuitry with their photonic counterparts. In optical routing and processing O RAM constitutes as a key device . This Paper provides an overview of various methods, materials, and structures used for constructing optical memory using photonic crystals by taking the advantage of strong confinement of photons and carriers and allowing heat to escape efficiently. This Paper compares the way for constructing a low-power large-scale o-RAM system that can handle high-bit-rate optical signals.

Keywords - *large-scale o-RAM,* *Optical Memory*. ***Photonic crystal***

I. INTRODUCTION

In today's Telecommunication scenario, photons have become the main carrier of information. In the earlier systems based on electronics, the data processing at nodes or routers was performed with integrated electronic circuits, this was comparatively slow, consuming high power, and was associated with heat generation during high bit rate operations. These were the real obstacles to further improvements in network speed and traffic capacity.

To improve the networking speed, optical data processing is preferred along with photonic integrated circuits to reduce the power consumption and the heat generation during high bit rate operations while keeping the high-speed properties of optical signals. The progress in optical Network communication, at present, is hindered by absence of an all Optical memory. Future optical routing and switching will require high-speed and low-power optical processing of digital signals[1,2]

PhC are periodic optical nano structures that affect the motion of photons and hence controlling and manipulating the flow of light. PhC dielectric function exhibits spatial periodicity in one, two or three dimensions. photonic crystal has several important advantages, such as small footprint, small energy consumption, high speed, and significant integratability. Thus it may be a promising candidate for all-optical information processing in optical networks.

N. SILICON CHIP BASED OPTICAL FLIP-FLOP MEMORY

The SOI circuit was fabricated with 193 nm DUV lithography using the ePIXfab silicon photonics platform. The SOI structure was designed with a top silicon layer of 220 nm and a buried oxide layer of 2 mm. An unpatterned III–V die was bonded on top of the finished SOI wafer/chip using DVS–BCB[3]. The III–V layer had a total thickness of 583 nm, including three compressively strained InAsP quantum wells for providing transverse electric mode gain and a tunnel junction for a low-loss p-contact. After removing the InP substrate, contact lithography was used to define the microdisk pattern to ensure that the underlying SOI waveguide was well aligned to the edge of the disk. The III–V layer was etched by inductively coupled plasma–reactive ion etching (ICP–RIE) until a thin n-doped InP lateral contact layer (90 nm) was achieved. This InP contact layer was then removed where it was not required, and a titanium/platinum/gold metal layer deposited on top to form the bottom contact. The whole structure was then covered with DVS–BCB. A via was opened through the DVS–BCB layer on the centre of the microdisk, and another titanium/platinum/gold metal layer deposited to form the top contact. The gold layer, which also served as a heat sink, was designed to be thick here (600 nm) to improve heat dissipation under continuous-wave bias. Because the WGM is confined to the edge of the disk, this top metal layer does not result in substantial optical absorption losses. The DVS–BCB was etched away on part of the bottom contact metal to enable it to be contacted electrically.

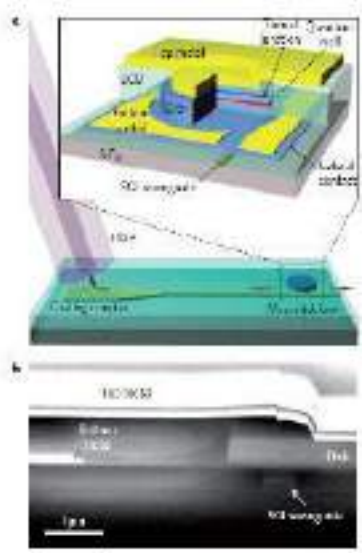


Figure1. Structure of the microdisk laser.

Result: a single microdisk laser with a diameter of 7.5 mm, coupled to a silicon-on-insulator wire waveguide, an all-optical flip-flop working in a continuous-wave regime with an electrical power consumption of a milliwatts, allowing

switching in 60ps with 1.8fJ optical energy. The total power consumption of 0.5 mw and the device size are 7.5_ 7.5 μm^2 [4] .

MM. ULTRA HIGH Q InGaAsP PHOTONIC NN.CRYSTAL BASED ON-CHIP OPTICAL BIT MEMORY

The figure below shows a scanning electron micrograph of PhC fabricated on an InGaAsP substrate by a combination of electron beam lithography and ICP dry etching. The air hole size and core thickness are both 200 nm. And adopted a width-tuned line defect cavity so that the air holes surrounding the cavity were shifted away from the center of the line defect.

The cavity mode volume is very small at 0.16 μm3 with a calculated Q-factor of over10 million, and realized a Q-factor of 1,800,000 in a Si-based PhC nanocavity. A light is inputted from a lensed fiber to a PhC-WG with a core of 1.05 W through an input/output waveguide with a 3 μm wide PhC line-defect and a spot size converter with a 15μm long elliptic arc. Here W is the basic line-defect width, which is defined as the distance between the center of adjacent holes of $a\ (3)^{1/2}$ and a is the lattice constant of the PhC.

.

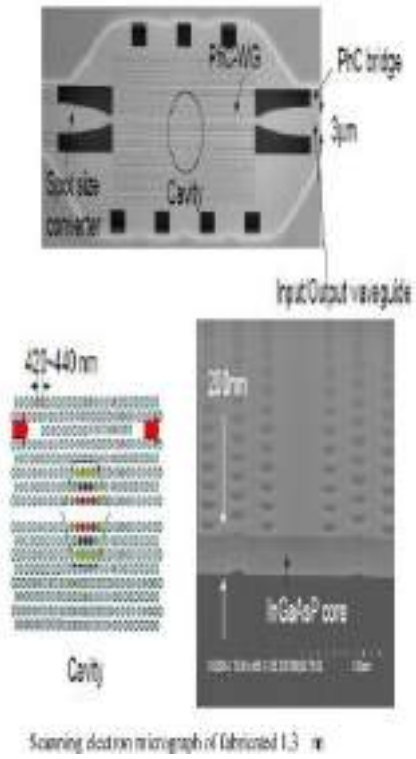


Figure 2. InGaAsP core PhC on InP substrate.

Result:In this method demonstrated all-optical bit memory operation with photonic crystal (PhC) nanocavities based on an InGaAsP substrate with a band gap at a wavelength of about 1.3 μm. The optical bistability is based on a refractive index modulation caused by carrier-plasma dispersion. The operating energy required for switching is only 30 fJ, and the minimum optical bias power for bistability is 40 μW, which is about one hundred times less than that required for laser based bistable memories[5].

IV ULTRALOW-POWER ALL-OPTICAL RAM BASED ON InP BASED InGaAsP BURIED HETROSRUCTURE

Device structure: Because the band-filling dispersion and linear absorption become strong in the vicinity of an electronic band-edge wavelength, and expect efficient carrier-induced nonlinearity. However, excess absorption degrades the cavity Q and increases the required operating power. Therefore, the composition of the buried InGaAsP should be adjusted to obtain both strong nonlinearity and moderate absorption. In the experiment, the composition was adjusted so that its electronic band-edge wavelength was 1.45 mm for an operation wavelength of 1.55 mm. The buried region was 4 mm long, 0.3 mm wide and 0.15 mm thick. The air hole diameter, lattice period and total thickness of the InP-based photonic crystal slab were 180, 425 and 250 nm, respectively. As shown in Fig. 1a, the photonic crystal waveguide width is increased to 1.1W0 for the guiding region and decreased to 0.98W0 for the barrier region, where W is the basic line defect width defined as the removal of one row of air holes in the G–K direction. The cavity region has a waveguide–width modulation with graded air hole shifts of 5, 10 and 15 nm.A nanocavity is formed by a combination of width modulation and index modulation induced by the buried InGaAsP.

For a 1-bit RAM measurement demonstration, a write pulse, a read pulse, a c.w. bias and a reset pulse are generated and used pulsed light from a tunable wavelength. Mode-locking fibre laser with a pulse width of 12 ps and a 10 MHz repetition rate, and a writing pulse with a 500 kHz repetition rate was picked out with a modulator. A read pulse was generated by splitting from the same laser, but the repetition rate was kept at 10 MHz. The read pulse was sufficiently attenuated that it did not influence the memory state. The c.w. bias light and reset pulse (width, 50 ns; repetition rate, 500 kHz) were obtained with a tunable-wavelength c.w. laser and an electro-absorption modulator. These light sources were merged using an optical coupler and set to a transverse electric polarization, defined as an electric field in the plane of the photonic crystal slab. These lights were then coupled by means of a lensed fibre into a 3-mm-wide photonic crystal waveguide and an intermediate spot size converter so that the light could be input efficiently into the single-line-defect photonic crystal waveguide[6]. The output lights from the sample were amplified by an erbium-doped fibre amplifier (EDFA), and the output waveform of the bias light and read pulse were selectively monitored with a band-pass filter and a sampling oscilloscope. In the measurement, there was an additional loss of 222 dB comprising the coupling loss between the optical fibre and the photonic crystal waveguide, as well as the propagation loss in the waveguide. Because the cavity is placed at the centre of the waveguide, the optical power injected into the cavity is estimated by measuring the power in the fibre and the insertion loss up to the cavityof 211 dB.

In this method demonstrated that photonic crystal nanocavities with an ultrasmall buried heterostructure design can solve most of the problems encountered in o-RAMs. By taking advantage of the strong confinement of photons and carriers and allowing heat to escape efficiently, and realized

all-optical RAMs with a power consumption of only 30 nW[6], which is more than 300 times lower than the previous record, and have achieved continuous operation. And also demonstrated their feasibility in multibit integration. This paves the way for constructing a low-power large-scale o-RAM system that can handle high-bit-rate optical signals.

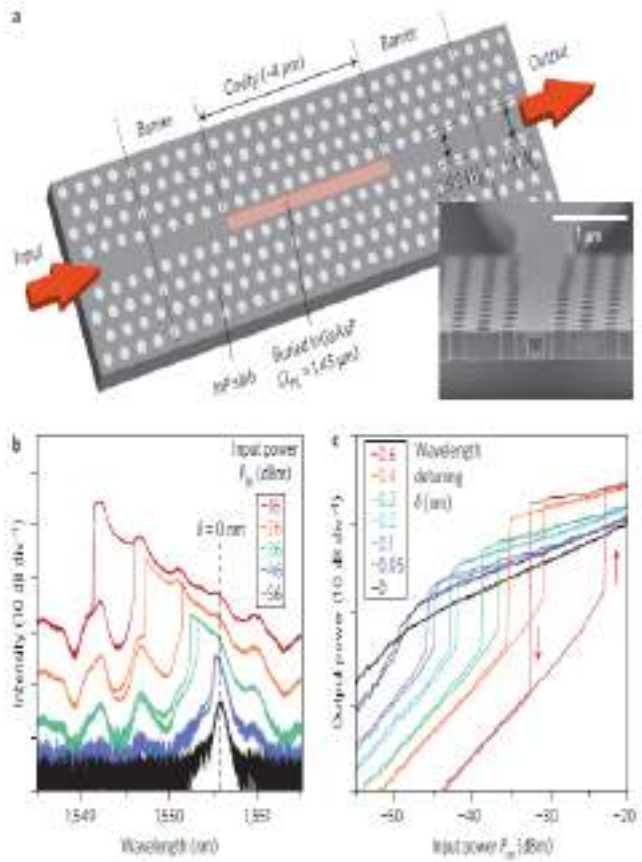


Figure3. Design of the BH–PhC nanocavity, The InGaAsP region
was etched to clearly reveal the BH region.

Summary

1 Mbit o-RAMs can be integrated in a footprint of the order of 10 mm^2, and the total power consumption remains less than 100 mW[6]. Although the number of bits is still smaller than that of electric RAMs, o-RAMs may play an important role as high-speed all-optical processing nodes where it is necessary to avoid E–O/O–E conversion.

If we integrate a certain number of bit memories, for certain applications, we may have to reduce the erasing time by sacrificing the carrier lifetime and subsequently the power consumption, but this will be acceptable because large-scale o-RAMs are not required.

The present all-optical bit memory based on a photonic crystal nanocavity is the most promising candidate for an o-RAM in the field of photonic processing circuitry, especially for all-optical packet routers for ultrahigh-speed photonic network processing.

Table 1 | Comparison of performances of various on-chip all-optical memories.

Device	Bias power (mW)	Switching energy (fJ)	Switching speed	Area (μm^2)	Ref.
MMI-BLD	3×10^3 to 5×10^3 (Elec.)	-	240 ps	7,000	8
VCSEL	4 (Elec.)	0.2	50 ps	36	7
Microdisk laser	6 (Elec.)	1.8	60 ps	45	9
Photonic crystal nanolaser	2.5×10^{-2} (Opt.)	-	On: 58 ps Off: 65 ps	<10	24
Photonic crystal nanocavity (all-InGaAsP)	1×10^{-2} (Opt.)	24	Off: 240 ps	<10	6,16
BH-PhC nanocavity	3×10^{-5} (Opt.)	2.5	On: 44 ps Off: ~7 ns	<10	This work

REFERENCES

W. Tucker, R. S. "Role of optics and electronics in high-capacity routers". IEEE J. Lightwave Technol. 24, 4655–4673 (2006).

X. Dorren, H. J., Calabretta, N. & Raz, O. "Scaling all-optical packet routers: how much buffering is required? J". Opt. Netw. 7, 936–946 (2008).

Y. Roelkens, G., Van Thourhout, D., Baets, R., Notzel, R. & Smit, M. "Laser emission and photodetection in an InP/InGaAsP layer integrated on and coupled to a silicon-on-insulator waveguide circuit". Opt. Express 14, 8154–8159 (2006).

Z. Liu Liu, Rajesh Kumar, Koen Huybrechts, Thijs Spuesens, Gu¨nther Roelkens, Erik-Jan Geluk, Tjibbe de Vries, Philippe Regreny, Dries Van Thourhout, Roel Baets and Geert Morthier "An ultra-small, low-power, all-optical flip-flop memory on a silicon chip" NATURE PHOTONICS 2010.

AA. Akihiko Shinya, Shinji Matsuo, Yosia1, Takasumi Tanabe, Eiichi Kuramochi, Tomonari Sato, Takaaki Kakitsuka, Masaya Notomi, "All-optical on-chip bit memory based on ultra high Q InGaAsP photonic crystal", OPTICS EXPRESS 19382 ,10 November 2008 / Vol. 16, No. 23 .

BB. Kengo Nozaki, Akihiko Shinya, Shinji Matsuo, Yasumasa Suzaki, Toru Segawa, Tomonari Sato, Yoshihiro Kawaguchi, Ryo Takahashi and Masaya Notomi, "Ultralow-power all-optical RAM based on nanocavities" NATURE PHOTONICS | VOL 6 | 248 APRIL 2012 .

[7] Kawaguchi, H., Mori, S., Sato, Y. & Yamayoshi, Y. "Optical buffer memory using polarization-bistable vertical-cavity surface-emitting lasers". Jpn J. Appl. Phys. 45, L894–L897 (2006).

[3] Park, H., Mack, J. P., Blumenthal, D. J. & Bowers, J. "An integrated recirculating optical buffer". Opt. Express 16, 11124–11131 (2008).

[4] Trita, A. et al. "Dynamic operation of all-optical flip-flop based on a monolithic semiconductor ring laser". European Conference on Optical Communication, paper We2C3 (2008).

[5] Mezosi, G., Strain, M. J., Furst, S.,Wang, Z. & Sorel, M. "Unidirectional bistability in AlGaInAs microring and microdisk semiconductor lasers". IEEE Photon. Technol. Lett. 21, 88–90 (2009).

[12] Mori, T., Yamayoshi, Y. & Kawaguchi, H. "Low-switching energy and highrepetition- frequency all-optical flip-flop operations of a bistable vertical-cavity

surface-emitting laser". Appl. Phys. Lett. 88, 101102 (2006).

[13] Roelkens, G., Van Thourhout, D., Baets, R., Notzel, R. & Smit, M. "Laser emission and photodetection in an InP/InGaAsP layer integrated on and coupled to a silicon-on-insulator waveguide circuit". Opt. Express 14, 8154–8159 (2006).

[14] Jalali, B. & Fathpour, S. "Silicon photonics". IEEE J. Lightwave Technol. 24, 4600–4615 (2006).

[15] Miller, D. A. B. "Device requirements for optical interconnects to silicon chips".Proc. IEEE 97, 1166–1185 (2009).

[16] Notomi, M. et al. "Optical bistable switching action of Si high-Q photonic-crystalnanocavities". Opt. Express 13, 2678–2687 (2005).

[17] Tanabe, T., Notomi, M., Mitsugi, S., Shinya, A. & Kuramochi, E. "Fast bistable all-optical switch and memory on a silicon photonic crystal on-chip". Opt. Lett. 30, 2575–2577 (2005).

[18] Soljac˘ic´, M., Ibanescu, M., Johnson, S., Fink, Y. & Joannopoulos, J. "Optimal bistable switching in nonlinear photonic crystals". Phys. Rev. E 66, 055601 (2002).

[19] hinya, A. et al. "All-optical on-chip bit memory based on ultra high Q InGaAsP photonic crystal". Opt. Express 16, 19382–19387 (2008).

[20] Liu, L. et al. "An ultra-small, low-power, all-optical flip-flop memory on a silicon chip". Nature Photon. 4, 182–187 (2010).

[21] Kitayama, K. et al. "Optical ram buffer for all-optical packet switches. Proceedings of the Communications and Photonics". Conference Exhibition (ACP2009), 1–2 (2009).

[22] Kuramochi, E. et al. "Ultrahigh-Q photonic crystal nanocavities realized by the local width modulation of a line defect". Appl. Phys. Lett. 88, 041112 (2006).

[23] Notomi, M. et al. "Low-power nanophotonic devices based on photonic crystals towards dense photonic network on chip". IET Circ. Device Syst. 5, 84–93 (2011).

[24] Tanabe, T., Taniyama, H. & Notomi, M. "Carrier diffusion and recombination in photonic crystal nanocavity optical switches". J. Lightwave Technol. 26,1396–1403 (2008).

[26] Nozaki, K. et al. "Sub-femtojoule all-optical switching using a photonic-crystal nanocavity".

[27] Nature Photon. 4, 477–483 (2010). Notomi, M. Manipulating light with strongly modulated photonic crystals". Rep.Progr. Phys. 73, 096501 (2010).

[28] Matsuo, S. et al. "High-speed ultracompact buried heterostructure photoniccrystallaser with 13 fJ of energy consumed per bit transmitted". Nature Photon. 4,648–654 (2010).

[29] Takahashi, R. et al. "Photonic random access memory for 40-Gb/s 16-b burst optical packets". IEEE Photon. Technol. Lett. 16, 1185–1187 (2004).

[30] Chen, C. H. et al. "All-optical memory based on injection-locking bistability in photonic crystal lasers". Opt. Express 19, 3387–3395 (2011).

[31] Baba, T. *et al.* "Light propagation characteristics of straight single line defect optical waveguides in aphotonic crystal slab fabricated into a silicon-on-insulator substrate". *IEEE J. Quant. Electron.***38,** 743–752 (2002).

[32] Kim, M. K., Hwang, I. K., Kim, S. H., Chang, H. J. & Lee, Y. H. "All-opticalbistable switching in curved microfiber-coupled photonic crystal resonators".Appl. Phys. Lett. 90, 161118 (2007).

[33] Weidner, E., Combrie, S., de Rossi, A., Tran, N. V. Q. & Cassette, S. "Nonlinear and bistable behavior of an ultrahigh-Q GaAs photonic crystal nanocavity". Appl.Phys. Lett. 90, 101118 (2007).

[34] Gibbs, H. M. "Optical Bistability: Controlling Light with Light" AcademicPress, 1985.

[35] Tanabe, T., Taniyama, H. & Notomi, M. "Carrier diffusion photonic crystal nanocavity optical switches". J. Lightwave Technol. 26, 1396–1403 (2008).

[36] M. Notomi, A. Shinya, S. Mitsugi, G. Kira, E. Kuramochi, and T. Tanabe, "Optical bistable switching action of Si high- *Q* photonic-crystal nanocavities," *Optics Express*, vol. 13, no. 7, pp. 2678–2687, 2005.

[37] T. Tanabe, M. Notomi, S. Mitsugi, A. Shinya, and E.Kuramochi, "Fast bistable all-optical switch and memory on a silicon photonic crystal on-chip," *Optics Letters*, vol. 30, no. 19, pp. 2575–2577, 2005.

[38] V. R. Almeida and M. Lipson, "Optical bistability on a silicon chip," *Optics Letters*, vol. 29, no. 20, pp. 2387–2389, 2004.

[39] M. Notomi, A. Shinya, S. Mitsugi, E. Kuramochi, and H.-Y.Ryu, "Waveguides, resonators and their coupled elements in photonic crystal slabs," *Optics Express*, vol. 12, no. 8, pp. 1551–1561, 2004.

Analysis of White Blood Cells for Malaria Detection

Hemalatha K L, Karthik R, Nandan Sateesh Hegde
Dept.of ISE, SKIT, Bangalore,
hema.sjcit@gmail.com, karthikrajendra4@gmail.com, nandan.hegde678@gmail.com

ABSTRACT— Malaria is a diseases caused by bacteria, viruses or parasites transmitted by mosquitoes. It is a serious global health problem. The current way of malaria parasite diagnosis is microscopic observation of the blood slide by a trained and experienced lab technician which takes more time and expensive due to the complex contents of the blood smear and the expertise requirement. This paper proposes a combination of color and morphological based segmentation techniques for blood cells segmentation on microscopic images. K-means clustering is implemented for color based clustering and primary segmentation of the image. The clustering is cascaded into two layers for cell recognition and background elimination. The proposed method overcomes the problem of previous work which is over segmentation and efficiently segments the WBC cells and keeps cytoplasm unaffected.

Keywords-blood cells segmentation; k-means clustering;

I. INTRODUCTION

Malaria is one of the predominant tropical diseases in the world causing wide spread sufferings and deaths in the developing countries[4]. In 2010, about 3.3 billion people which are half of the world populations are at risk of malaria. In addition, this disease has caused the death of an estimation of 655,000 people in 2010, with 86% of the victims are children under five years of age [1]. The proportion of counts of different types of white blood cells in the bone marrow, called differential counts, provides invaluable information to doctors for diagnosis. Due to the tedious nature of the differential white blood cell counting process, an automatic system is preferable [2].

The need for automation of leukemia detection arises since current methods involve manual examination of the blood smear as the first step toward diagnosis. This is time-consuming, and its accuracy depends on the operator's ability [3]. The clinical behavior of the disease can be predicted using this classification and accordingly treatment should be given to the patient. In leukemia disease, large numbers of abnormal white blood cells are produced by bone marrow due to unknown cause. In pathology manual detection of leukemia is done which is time consuming as well as costly due to high
cost pathology instruments. Hence automatic technique is adopted for fast and accurate results [5].
This paper proposes a combination of color and morphological based segmentation methods for blood cells segmentation on microscopic bone marrow images. The main goal of this paper is to reduce over segmentation phenomenon by using k-means algorithm. Kmeans clustering is the color based technique performed as a primary segmentation stage to cluster the image pixels according to the similarity of color components and spatial coordinates.

One of the disadvantages of diagnosis using manual microscopy methods is that it requires extensive human intervention during the diagnostic process which can often lead to late and sometimes erroneous diagnosis. The microscopist requires extensive training to gain expertise in the diagnosis, and because of the sheer volume of the samples that need to be analysed, the method is not consistent and is dependent upon blood smear and stain quality, microscope quality and the expertise of the microscopist. [7]

Some of the problems of manual microscopy can be overcome by exploring computer based, specifically image-based, diagnostic methods. The aim of this study is to outline a semi-automatic diagnosis method based on image processing and one that provides a reliable and consistent solution. The literature contains descriptions and details of several computer vision or image-based algorithms [8].

As malaria is a health care issue and is causing a serious health problem, the detection in the early stage is very important.

II. METHODOLOGY

The proposed procedures to develop a model for the Analysis of white blood cells for malaria detection overall working principle is depicted in Fig. 1. The steps are summarized as follows:

Step 1: Read the malaria slide images as the input.

Step 2: (Red Green and Blue) RGB to Grayscale conversion.

Step 3: Image filtering and thresholding.

Step 4: Erosion and Dilation

Step 5: Apply K-means for image segmentation.

Step 6: Labeling and counting.

STEP 1: Read the malaria slide images as the input:

In this step, the stained blood slide is captured and is given as an input to the System. The images are usually collected from the labs and the Hospitals for detection of malaria.

Step 2: RGB to Grayscale conversion:

Once the image is read into the system, it is converted to a Grayscale image. The contrast of the gray scale image is enhanced using local histogram equalization to enhance the visibility of the parasites and RBC.

Step 3: Image filtering and thresholding:

Filtering is done to reduce noise within an image or to produce a less pixilated image. Most filtering methods are based on low pass filters. There might be some unwanted regions or noises that are still encountered in the image. Thus, median filter is used as a noise removal in order to obtain a noise-free image

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. The input to a thresholding operation is typically a grayscale or color image. In the simplest implementation, the output is a binary image representing the segmentation. Black pixels correspond to background and white pixels correspond to foreground (or *vice versa*). In simple implementations, the segmentation is determined by a single parameter known as the *intensity threshold.* In a single pass, each pixel in the image is compared with this threshold. If the pixel's intensity is higher than the threshold, the pixel is set to, say, white in the output. If it is less than the threshold, it is set to black.

Step 4: Erosion and Dilation:

The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the *structuring element* used to process the image. In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. The rule used to process the pixels defines the operation as dilation or erosion.

Dilation is one of the two basic operators in the area of mathematical morphology, the other being erosion. It is typically applied to binary images, but there are versions that work on grayscale images. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels. Thus areas of foreground pixels grow in size while holes within those regions become smaller.

Erosion is the other operator of mathematical morphology, the basic effect of the operator on a binary image is to erode away the boundaries of regions of foreground pixels (*i.e.* white pixels, typically). Thus areas of foreground pixels shrink in size, and holes within those areas become larger.

Step 5: Apply K-means for image segmentation:

After transforming the RGB into grayscale, the next and important step in image segmentation is to extract the meaningful region from malaria image.

Clustering is a method to divide a set of data into a specific number of groups. It's one of the popular method is k-means clustering. In k-means clustering, it partitions a collection of data into a k number group of data [6]. It classifies a given set of data into k number of disjoint cluster. K-means algorithm consists of two separate phases. In the first phase it calculates the k centroid and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance.

Once the grouping is done it recalculate the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by its member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. So K-means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over all clusters.

K-Means algorithm:

K-means clustering[9] is an iterative, data-partitioning algorithm that assigns n observations to exactly one of k clusters defined by centroid, where k is chosen before the algorithm starts.
The algorithm proceeds as follows:

1. Choose k initial cluster centers (centroid)
2. Compute point-to-cluster-centroid distances of all observations to each centroid.
3. There are two ways to proceed :
   - **Batch update** — Assign each observation to the cluster with the closest centroid.
   - **Online update** — Individually assign observations to a different centroid if the reassignment decreases the sum of the within-cluster, sum-of-squares point-to-cluster-centroid distances.
4. Compute the average of the observations in each cluster to obtain k new centroid locations.
5. Repeat steps 2 through 4 until cluster assignments do not change, or the maximum number of iterations is reached.

Step 6: Labeling and counting:

Based on the K-means clustering algorithm, the cells that are affected by malaria can be detected. We can label the cells that are affected by malaria and also the total number of cells that have been tested. The number of blood images affected by malaria. We can label the cells affected by malaria and a count of cells affected.

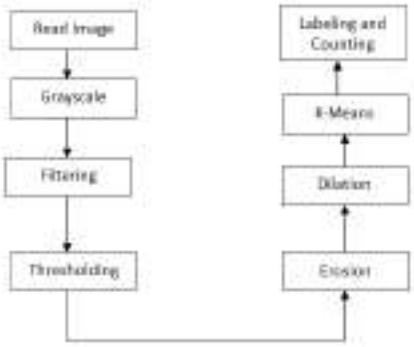


Figure 1
Block diagram of the proposed methodology

I. RESULT

The proposed technique has been applied on few peripheral blood smear images obtained from the public dataset. A microscopic blood image of size 700 × 468 and jpeg (figure 2) format is considered for evaluation. The image is first converted into a grayscale image as shown in figure 3. Once the image is converted into the Grayscale image, the image is clustered. The clustered image is shown in figure 4. The image is then segmented (figure 5). The result also contains the cells affected by Malaria parasite. Figure 6 shows the screen shot of the final result, the total number of cells tested were 23 and the affected cells were 4

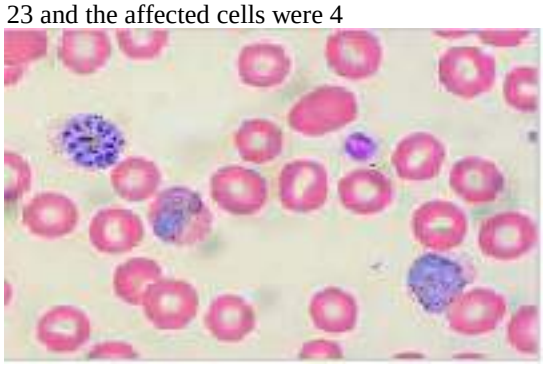


Figure 2: Input Cell Image

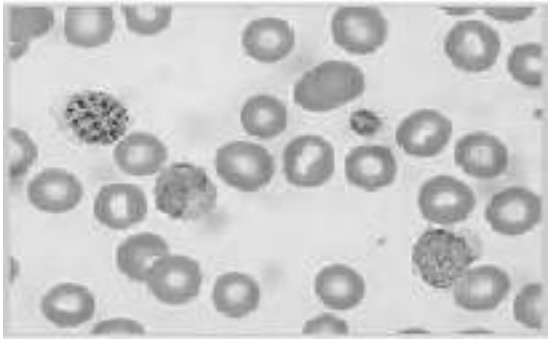


Figure 3: Grayscale Image

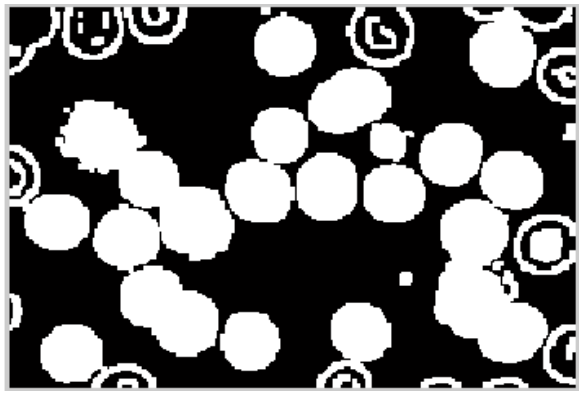


Figure 4: Clustered Cell Image

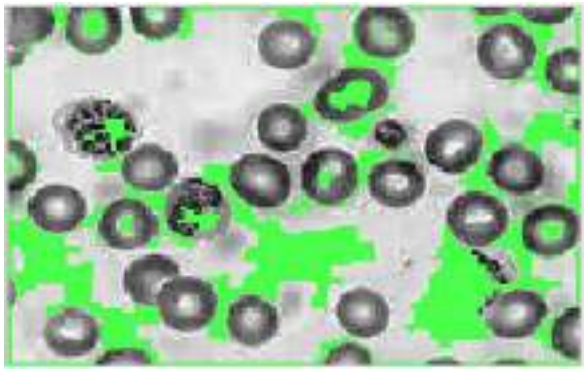


Figure 5: Segmented Cell Image

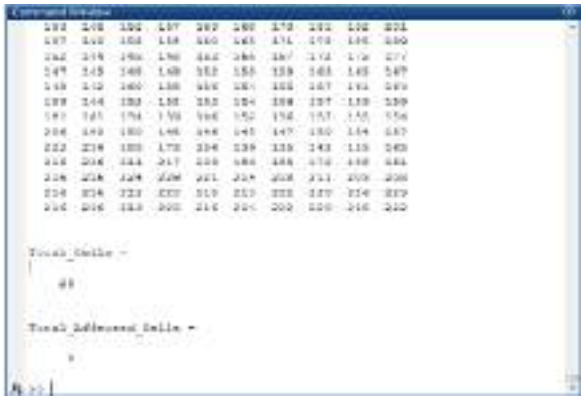


Figure 6: Screen shot of the final output

II. CONCLUSION

The image segmentation is done using k-clustering algorithm, by initially creating the k centroid. We have made the erosion and Dilation which improves the quality of original image. The filtering of the image is done using the median filter. Hence, we can conclude that the proposed clustering algorithm has better accuracy and is more robust in nature.

III. FUTURE WORK

The methodology can be improved to obtain more efficient and quicker results. The k-means algorithm can be modified with the combination of other algorithms to obtain better system. Also, we can improve the quality of the output

image more by using the morphological operation and get better performance measurement. We can also implement different clustering method using subtractive clustering algorithm and lastly we can implement and analyze in different areas of image segmentation. The quality of the image can be improved and the system can be made more accurate with better efficiency.

IV. REFERENCES

[1] WHO, World Malaria Report 2011, World Health Organization, 2011.

[2] N. Theera-Umpon, S. Dhompongsa (2007). Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification, in Information Technology in Biomedicine, IEEE Transactions on, 2007.11(3): pp 353-359.

[3] S. Agaian, M. Madhukar, A.T. Chronopoulos (2014). Automated screening system for acute myelogenous leukemia detection in blood microscopic images, in IEEE Systems Journal.

[4] The Malaria Control Programme, World Malaria report, World Health Organization, 2008

[5] H.N. Lim, M. Mashor, R. Hassan (2012). White blood cell segmentation for acute leukemia bone marrow images, in Journal of Medical Imaging and Health Informatics Volume 2, Issue-3, May–June 2013.

[6] R. Yager and D. Filev, Generation of Fuzzy Rules in Mountain Clustering, In Journal of Intelligent and Fuzzy System, vol. 2(3), pp. 209–219, (1992).

[7] Yashasvi Purwar, Sirish L Shah, Gwen Clarke, Areej Almugairi, and Atis Muehlenbachs (2013). Automated and unsupervised detection of malarial parasites in microscopic images, in Journal of Malar J. Volume 10, Dec 2013.

[8] Rao KNRM, Dempster AG, Jarra B, Khan S. Automatic scanning of malaria infected blood slide images using mathematical morphology. Proc IEE Semin Med Appl of Signal Process, London, UK.2002.

[9] Lloyd, Stuart P. "Least Squares Quantization in PCM." IEEE Transactions on Information Theory. Vol. 28, 1982, pp. 129–137.

3D Reconstruction Methodologies: a Review

Sowmya S G, Dr. Sunanda Dixit
Dept. of Information Science and Engineering, DSCE, Bangalore, India
sowmya.sg1791@gmail.com, sunanda.bms@gmail.com

ABSTRACT -Usually underwater images will have a less visibility conditions, for which the 3D reconstruction for these images has become a challenging tasks. This is because of problems like harsh environmental conditions (black smoke, white smoke) or may be of infraction by floating partials and aquatic animals. The other problem that arises like light propagation in underwater which affects the image color, and the problem in visualization of small objects is poor.to overcome these problems in this paper, we did the detailed survey on different technique of 3D reconstruction of underwater objects.

Keywords- SFM, Speeded Up Robust Feature, CLAHE, SIFT, Delaunay Triangulation.

I. INTRODUCTION

In computer vision, 3D object reconstruction is a process to recreate the shapes and appearances of original objects. 3D reconstruction can be performed by Active method and Passive method. The active method reconstructs the 3D objects either by using a mechanical method or a radiometric method. Passive method don't interfere with the reconstruction technique, as it uses sensors to estimate the radiance released by surface of objects to infer its 3D modal.

After the revolutionary contributions in medical practices and many other fields, 3D imaging system is used in more and more applications such as video surveillances, in the survey of underwater objects like biological species, ancient shipwrecks and coral reefs etc., video simulation and mapping. Now a days underwater imaging (marine science and technology) has become a very interesting topic for researchers.

Underwater imaging will have a poor visibility conditions. This is because of challenging environmental conditions (black smoke, white smoke) or may be violated by fish, floating particles and aquatic animals. Another challenge that arises on image color when light propagation occurs in underwater. Another challenge is the visibility of small objects is not good or hidden when the underwater image is captured. To overcome these problems in this paper, we did the detailed survey on different technique of 3D reconstruction of underwater object.

II. LITERATURE OVERVIEW

Luca Mazzei et al.[1] the proposed work presents an autonomous, small sized, low cost device for underwater imaging. The low cost system is used to analyze marine organism (macro, mega zooplankton and benthos) through reconstructing 3D models. The system is designed completely autonomous, low cost and small in size, in order to deploy easily on any underwater vehicle and is also used in the fixed area as a stand-alone mode to monitor the benthos growth for long time, and because of its flexibility it is used in many underwater investigation approach even in harsh environment. The system is capable of reaching deep-sea depth because of its autonomous functionality, without the support of any divers. The image acquisition task is completely autonomous and it is designed to be fully programmable in camera parameter and performs in real time. Then the 3D reconstruction is performed by using metadata and the information collected during the image acquisition process.

Adrian Bodenmann et al.[2] the proposed work demonstrate that the underwater imaging system can generate the high resolution 3D color reconstruction of the seafloor, from the range of up to 13m. Visual mapping is used in the applications like survey of underwater archeology, survey of marine organism, to inspect man-made structures etc. Usually the attenuation of light in water is more than in air, so to obtain the colored underwater images it is required to be within 2-3m from the seafloor. In such low altitude, the underwater vehicle can acquire a limited area in each image of the seafloor, which is time consuming for the survey of large area. To overcome this problem, 3D mapping device "SeaXerocks" has been developed to increase the efficiency of acquiring large area of the seafloor images, also obtain high quality of color images from the high altitude of up to 13m. The algorithm uses these images to automatically process the data it obtains.

Dr. Ramakanth Kumar P et al.[3] proposed a 3D object reconstruction system using a feature detection technique called Speeded Up Robust Feature(SURF) is a robust feature detector and is used in 3D reconstruction problem and object recognition problem. Speeded up robust feature points and are matched between the pair of underwater images. The unwanted feature points are removed using epipolar geometry and derive the geometrical relation. Then the 3D points are computed using linear triangulation. These 3D point clouds are then used for 3D reconstruction and texture mapping.

Pulung Nurtantio Andono et al.[4] proposed a Contrast Limited Adaptive Histogram Equalization(CLAHE) enhancement technique for 3D surface reconstruction of underwater coral reef. Underwater images will have a less quality of visibility because of absorption and scattering of light, which affects the image color. To enhance the quality of the images for underwater image CLAHE algorithm is used in preprocessing area. The image pairs of the seafloor were manually extracted from the video footages, to develop a 3D representation of a seafloor. The SIFT (Scale Invariant Feature Transform) image matching algorithm is used to automatically extract the corresponding points between the image pairs. The performance of the Scale Invariant Feature Transform with CLAHE is computed using the number of matching points. Thus the combination of CLAHE and outlier removal provides a better enhancement technique to improve the quality of the image.

Alessandro Gallo et al.[5] proposed work presents the analysis made on small specimens, using a multiview stereo technique to reconstruct the 3D object. The analysis in the field of cultural heritage is gaining importance for

restoration, duplication and maintenance of ancient specimens. The 3D reconstruction technique is used to determine the characteristics of a particular species. The problem rose due to use of macro lenses, such as small depth and loss of sharpness because of diffraction. To overcome this problem, every image in the sequence is acquired by merging all the images captured at different focus plane using image fusion algorithm. This results in obtaining the object with high quality textured 3D modal that ranges from some millimeter to centimeter.

Prabhakar C J et al.[6] the proposed work presents the reconstruction of 3D surface for the underwater objects. The surface reconstruction of 3D underwater image is a very challenging task, because of the poor visibility of underwater image. There are many reasons for the poor visibility of underwater image such as illumination of light propagation, scattering and floating particles that produce noise in the captured image. The preprocessing is applied for the degraded images by using wavelet denoising, anisotropic filtering and homomorphic. To rectify the preprocessed image uncalibrated rectification technique is applied. Then the graph cut method is used to find the corresponding points from the rectified image, which is the used to estimate depth of the image by applying triangulation method.

Pulung Nurtiantio Andono et al.[7] the proposed work presents the 3D reconstruction of coral reef images using lowcost underwater camera. A footages of the seafloor is collected from the multiview underwater cameras in linear transects. The image pairs of the seafloor were manually extracted from the video footages, to develop a 3D representation of a seafloor. The SIFT (Scale Invariant Feature Transform) image matching algorithm is used to automatically extract the corresponding points between the image pairs. Based on the result obtained by the corresponding points of the image pair the 3D position of the coral reef can be determined by using triangulation technique and the surface reconstruction is processed using Delaunay triangulation.

Georgios Papadopoulos et al.[8] proposed, SCOUT autonomous surface vehicle which is used to determine the above part and below part of the underwater structure. The autonomous surface vehicle is designed which is equipped with sonar and laser scanner, which is capable of scanning the above and below part of the underwater structure. To construct 3D models of the above part, scan matching technique is used. The 2D sonar data of the below part is then used by the transformation of the above part to construct a 3D modal. Then the above and below part of the 3D modal is combined to construct an entire 3D modal of the partially immersed underwater structure. Two types of maps are constructed: low quality map- constructed online which is used for navigation, and high quality map-constructed offline which is used for inspection.

Chris Beall et al.[9] the proposed work presents the 3D reconstruction of dense objects using smoothing and mapping techniques. There are technique which gives high quality 3D reconstruction for large structure, but the 3D reconstruction for small underwater objects is not satisfied because of the challenging environment. The dense object modal is built from the synchronized high definition video. The SURF is used to detect features, the camera trajectory and the 3D points are estimated, and are fed as the input to mapping and smoothing techniques for optimization. These 3D modals provides accuracy for quantitative measurements of drown structures.

Anne Sedlazeck et al.[10] proposed SFM (structure from motion) system allowing to compute detailed 3D reconstructions from underwater objects or scenes. The underwater video of deep seafloor structure from ROV Kiel 6000 is used for 3D reconstruction. The SFM (structure from motion) is used to provide a good imaging condition, because of which the camera is not required to follow the specialized trajectory. It adapts a special filtering to remove the noisy background and floating particles from the underwater images, which in turn is used to predict a camera pose and spare 3D points. Based on the camera path estimation, the dense depth is computed for each image, which enables to generate 3D surface modal. Once the 3D surface modal is generated, color correction is made by using physical modal for underwater light propagation, this allows to view the object without the effect of attenuation and scattering due to water.

V Brandou et al.[11] proposed work is to improve the 3D reconstruction method for small scale by image acquisition method for quantitative measurements. The camera parameter should be known, from which the image sequences are captured, and is used to compute the 3D modal for metric measure. The acquisition method is used to compute the extrinsic camera parameter by using camera trajectory. The procedure of 3D reconstruction is reduced and improved by capturing images in regular intervals. 3D reconstruction of the objects is performed using different techniques, then the texture map is applied to the dense 3D modal.

Oscar Pizarro et al.[12] proposed work presents when camera moves over a non-planer surface, recovering 3D structure to provide a global view for an area of interest, this is a local to global approach. This can be achieved by applying structure from motion with some additional requirements. These techniques are complex than mosaicking. Navigation sensors are used in vehicles which is used to generate 3D sub map, which in turn used for bundle adjustment for 3D structure.

Nathalie Pessel Ifremer et al.[13] proposed a self-calibration technique for a camera mounted on an underwater vehicle, that is used to perform the 3D reconstruction of underwater scenes. Underwater vision is a necessary element in subsea operations. The self-calibration technique is used for a camera mounted in a fixed position on an underwater vehicle. The main goal is to identify the intrinsic parameters of the camera. The navigation sensors are used to measure the motion of the vehicle, which in turn is used to estimate the extrinsic parameter. The moving camera will not affect the applications that uses stereoscopic method, nevertheless it enables robust algorithm that is used for point matching. Robust parameter is used to estimate the analysis of the number of point matches accurately.

III. CONCLUSION

3D reconstruction helps to recover correct shape and size of the objects. It helps to enhance the quality of underwater

image. The extensive literature survey shows many methods are existing for 3D underwater reconstruction. The different methodologies, advantages and disadvantages have been severed and listed. This helps to implement 3D reconstruction technique efficiently. This paper gives detail survey of algorithms used for 3D reconstruction of underwater images and their advantages and disadvantages.

REFERENCES

[1] Luca Mazzei, Lorenzo Corgnati, Simone Marini, "Low Cost Stereo System for Imaging and 3D Reconstruction of Underwater Organism", *2015 IEEE.*

[2] Adrian Bodenmann, Blair Thorntonand Tamaki Ura, "Development Of Long Range Color Imaging For Wide Area 3D Reconstruction Of The Seafloor", *2013 IEEE.*

[3] Rashmi Gokhale, Dr. Ramakanth Kumar P, "Analysis of Three Dimensional Object Reconstruction Method for Underwater Images", *International Journal of Scientists & Technological Research, volume 2, June 2013.*

[4] Pulung Nurtantio Andono, Riardus Anggi Puramunendar, Catur Supriyanto, Guruh Fajar Shidik, I Ketut Eddy Paranama, Mochamad Hariadi, "Enhancement Of 3D Surface Reconstruction Of Underwater Coral Reef Base On SIFT Image Matching Using Contrast Limited Adaptive Histogram Equalization and Outlier Removal", *Kursor Journal, vol.7, no.1, January 2013.*

[5] Alessandro Gallo, Fabio Bruno, Maurizio Muzzupappa, "Multi-View 3D Reconstruction of Small Stone Samples Deteriorated By Marine Organisms", *2012 IEEE.*

[6] Prabhakar C J, Praveen Kumar P U, "3D Surface Reconstruction of Underwater Objects", *International Journal of Computer Applications, April 2012.*

[7] Plung Nurtantio Andono, Eko Mulyanto Yuniarno, Mochamad Hariadj, Valentijn Venus, "3D Reconstruction Of Underwater Coral Reef Images Using Low Cost Multi-View Cameras", *2012 IEEE.*

[8] Georgios Papadopoulos, Hanna Kurniawati, Ahmed Shafeeq Bin Mohd Shariff, Liang Jie Wong, Nicholas M Patrikalakis, "3D surface Reconstruction For Partially Submerged Marine Structure Using An Autonomous Surface Vehicle", *2011 IEEE.*

[9] Chris Beall, Brain J Lawrence, Viorelalla, Frank Dellaert, "3D Reconstruction of Underwater Structure", *2010 IEEE/ RSJ International Conference on Intelligent Robots and Systems.*

[10] Anne Sedlazeck, Kevin Koser, Reinhard Koch, "3D Reconstruction Based On Underwater Video From Kiel 6000 Considering Underwater Imaging Conditions", *2009 IEEE.*

[11] V Brandou, A G Allais, M Perrier, E Malis, P Rives, J Sarrazin, P M Sarradin, "3D Reconstruction Of Natural Underwater Scenes Using The Stereovision System IRIS", *2007 IEEE.*

Automatic road extraction from satellite image by Difference of Gaussian and convolution overlap add method

Veda.D, Suma B Rao
[1]PG Scholar, [2]Assistant professor, Department of Computer Science, RVCE, Bangalore.
vedadgowda@gmail.com, sumab_rao@rvce.edu.in

ABSTRACT-The road network detection is an emerging area in information extraction from high-resolution satellite images (HRSI). The proposed work in this paper consists of three steps: Pre-filtering, connected component-based operations, and removal of unwanted non-road pixels using morphological operations. The proposed algorithm is experimented on various satellite images and the performance of the system is evaluated by comparing the results with ground truth road map as reference data, and performance measures such as completeness, correctness, and quality are calculated.

Keywords – **Gaussian Filtering, Gradient operation, Convolution, Smoothening, Edge Detection, Morphological operations.**

1. INTRODUCTION

Automatic road detection is particularly employed in the city planning, cartography and to revise already detected roads in Geographic Information Systems environment. Manual manipulation of GIS database is costly, time-consuming process, and also there is a possibility of error. Therefore, automatic road feature extraction from high resolution satellite image is required to detect the road network in a robust manner. The process of road detection from remote sensing images is quite complex, due to the presence of various noises. These noises could be the vehicles, crossing lines and toll bridges. Few small and large false road segments interrupt the extraction of road segments that happens due to the similar spectral behavior in heterogeneous objects.Images obtained from satellite are useful in much environment application such as tracking of earth resources, geographical mapping, and production of agricultural crops, urban growth, weather, flood and fire control.

The objective of road feature extraction method is providing a binary mask in which true pixels represent road regions, and false pixels indicate non-road regions. The major problem of road extraction method is the complex structure and texture of the images, which contain many different objects, such as roads, houses, trees, vehicles etc., with differences in shape, tone.

Road extraction methods can be classified into two types such as semi-automatic and fully automatic. The road detection methods which requires human interaction are known as semi-automatic, and those that are not requires human interaction are known as automatic and semi-automatic are not suitable for real-time application.

2. RELATED WORK

Sujatha et al. [1] has proposed an automatic road extraction system which uses connection component-based techniques for the high resolution satellite images. The algorithm is implemented on different satellite images and the results are evaluated for correctness, completeness and quality with true road maps as reference.

Jiuxiang[2] as proposed an automatic road detection algorithm which will include three stages such as identify the road pixels, track the road and road segments are grown. A road tree is constructed in order to identify the network of the roads.

Xiaoying et al. [3] used the combination of two detectors: segmentation based and fuzzy based detectors in order to extract the centerlines. The outputs of these two detectors are then combined to extract the road network. He used path searching algorithms to integrate the outputs.

Cem et al. has also proposed a method based on the probability and graph theory. In this method, first the edge pixels are extracted and using them the road centerline is obtained. At the last stage, the graph theory is implemented which will represent the extracted road regions in the form of graph. This system completely depends on the edge detection of the image [4].One more road extraction algorithm proposed by **Zhujian et al. [5]** which are based on the two methods. This system cross-validates the features of the Line Segment Detector (LSD) and Statistical Region Merge (SRM) based on their spatial relationship.

Xiangguo in his proposal classified the road into salient and non-salient markings and then used various tracking systems to detect roads. The salient road markings are extracted using the least squares interlaces matching; match the profiles, rectangular matching and texture signatures. All these are combined to extract the road network with non-salient markings [6].

Senthilnath et al. [7] proposed an automatic road detection system which can be used in urban areas. This system is based on the road characteristics like structure, geometric and spectral. The main process of this system is to reduce the non-road unwanted regions and used texture progressing analysis and cut methods to extract the road segments. The result is compared with the true road map in order to evaluate the performance measures.

Support vector machine (SVM) based road centerline extraction is proposed by Xin et al. [8]. In this algorithm, both the geometric and spectral characteristics are used. The hybrid combinations of the spectral and geometric data are analyses using the SVM analyzer. Hang et al. [9] has

also proposed a hybrid method for road extraction for both urban and rural regions. Gabor filtering is used along with the threshold based histogram to mark the surface and lane segments on the rural images. SVM analyzer is used in order to extract the road surface in the urban area. For further enhancement Gabor filtering is used. The experimental results were impressive.

Sukhendu et al. [10] has presented an automatic method for road extraction which is based on pre-dominant features of the road. Probabilistic SVM analyzers and Dominant Singular Measures are used for the segmentation of the road areas. Numerous post processing are done in order to remove the non-road regions from the image. This actually helped to solve the problems with the discontinuity.

Tieling et al. [11] proposed a road detection system which used the wavelet transform from the high resolution satellite images. In this work there are two parts. In the first part, the road edges are determined using the two dimensional wavelet transform. In the next part, thinning algorithms are performed which will detect the thin centerline in the road networks.

Sahar et al. has proposed an algorithm on satellite images which uses both particle filtering and extended Kalman filtering[12]. These two filtering mechanisms are applied in order to trace the roads which lies beyond the obstacles and also to track the different branches of road once the junction is found. To evaluate the results, the output is compared with the manually drawn road maps.

Thierry et al. [13] has proposed an integrated approach for road segmentation. This method consists of three stages. In the first stage, the non-road regions in the image are removed by using the watershed transform filters. Then the closing operation is done to extract the structure of the road network, . In the next stage, the graph is built which shows the relationship between the watershed lines. Finally, Markov random field is defined on that graph which will extract the road networks in the satellite images. Mena et al. [14] proposed an automatic road detection method for the rural and semi-urban regions. In this method, there are four modules such as pre-processing module, binary segmentation based on texture progressive analysis, vectorization of the binary image by means of skeletal extraction and morphological operations, and performance measures evaluation by comparing with the true maps.

Boshir et al. [15] has implemented an automatic road extraction method which uses road intersection model. The process of detecting the road intersection is done in two modules. Initially, road network is detected using various morphological filters and then the road intersection are extracted to check on the orientation of the roads. This algorithm is implemented on different images to evaluate the correctness of the method.

Jun Zhou in his paper [16] has proposed a system which updates the roads in the map be comparing the aerial images with the map to detect the new roads. Their method could not be fully automated as these algorithms are not reliable. Human interactions are needed for the final check of the maps.

3. METHODOLOGY OF PROPOSED WORK

The detailed methodology of proposed work is given in Figure 1. This includes pre-processing, pre-filtering, gradient operation, morphological operation and coloring.

3.1 Pre –Processing

Initially we convert the given satellite image SI into a Gray scale image $\mathbf{SI_G}$. The gray scale image many consists of small objects. So we remove the small objects by applying the morphological operation on gray scale image such as erosion and dilation. Morphological operation are mainly used for extracting image components used for future extraction such as boundaries, skeleton and convex hull.

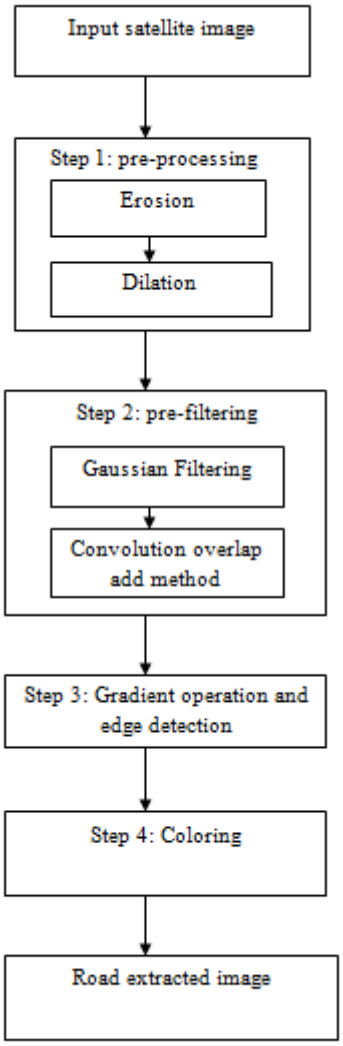


Figure 1: Methodology of proposed work

Dilation and erosion are the fundamental tools of morphological image processing. Dilation operation grows or thickens the objects in gray scale image. Let $\mathbf{SI}_G$ be an image, and B is the structuring element, then the dilation is represented as follows

$$SIg \oplus B = \{Z|\ (B^)_z \cap y = \phi\}.....(1)$$

Morphological erosion shrinks or thins objects in a binary image. The extent of this shrinking is controlled by structuring element. The erosion of dilated image SI_D by B is denoted as follows.

$SI_D \Theta B = \{Z|(B)_z \subseteq Y\}$ .....(2)

3.2 Pre –filtering

The pre-filtering process is used for removing certain types of noise from the satellite image. For pre-filtering the following algorithms is used.

Here we propose a pre-filtering process based on the **Difference of Gaussian** (DoG) as in equation 3.Then we merge the images using convolution overlap add method. For the erode image SI_E we apply **DoG** as follows.

$$\Gamma_{\sigma 1 \sigma 2}(x,y) = SI_E * \frac{1}{2\pi\sigma 2} e^{-(x^2+y^2)/(2\sigma^2)} - \frac{1}{2\pi\sigma 2} e^{(x^2+y^2)/(2\sigma^2)} .....(3)$$

```
Convolution(img1,img2)
    {
      n1←length of img1
      n2←length of img2
     compute n←n1+n2-1
     y←array of zeroes(1,n)
     h1←(dog zeros(1,n2-1)
     n3←lemgth of h1
     y←array of zeros(1,n+n3-n2)
     h←Fourier Transform(h1)
      for i=1 to n2
          if i<=(n1+n2-1)
            compute x1 =append zero
            array of size (1,n3-n2)to img1 of size(1,n3-n2)
        else
              compute x1=append zero
            array of size (1,n3-n2)to img1 (n1)
        end
        x2=fft(x1);
        x=x2*img2;
        x4=round(ifft(x3));
         if(i==1)
            normalize y(1:n3)=x4(1:n3)
         else
           normalize y(i:i+n3-1)=y(i:i+n3-1)+x4(1:n3)
        end
     end
  }
```

In image processing, **DoG** is an advance enhancement algorithm that involves the subtraction of one blurred version of an original image from another. In the simple case of grayscale images, the blurred images are obtained by convolving the original grayscale images with Gaussian kernels having differing standard deviations. Blurring an image using a Gaussian kernel suppresses only high-frequency spatial information. Subtracting one image from the other preserves spatial information that lies between the ranges of frequencies that are preserved in the two blurred images. Thus, the **DoG** is a band-pass filter that discards all random high frequency noise that is present in the original grayscale image.

After performing **DoG** we apply convolution by using **Convolution overlap add method.** It is a process of merging two images that is gray scale image SI_G and image obtained after **DoG** according the algorithm as shown below.

3.3 Gradient Smoothening

i. For the convoluted image, perform the Image Smoothening using the gradient minimization method.
ii. For the smoothened image, use Canny edge detection technique. This will sharpen the road boundaries.
iii. Perform the morphological filling on the edge detected image in order to remove any noises within the road regions.
iv. Use area based noise removal which will remove the non-road region noises.
v. Apply morphological dilation on the binary image to thicken the road objects.

3.4 Coloring

Once the binary image is dilated, the road region pixels are acquired. These are the pixels which will fall under the road regions. In the original image, the respective pixel values are changed to any color, so that the output image will have the road regions in different color.

4. PERFORMANCE EVALUATION OF ROAD EXTRACTION.

The performance of the proposed road extraction of observed by calculating the performance measures such as completeness, correctness and quality.

The matched extracted road data are calculated as true positive(TP).and unmatched extracted data is calculated as false positive(FP).The unmatched reference data are calculated as false negative(FN).

4.1 Completeness

It is the ratio of matched reference road data with total length of reference road map as shown in equation 4

$$\text{Completeness } \% = \frac{TP}{TP+FN} * 100 \quad \text{..... (4)}$$

4.2 Correctness

It gives the percentage of correctly extracted road i.e. ratio of matched parts of extracted road with total length of extracted road as shown in equation 5

$$\text{Correctness \%} = \frac{TP}{TP+FN} * 100 \quad \textbf{..... (5)}$$

4.5 Quality

It considered both completeness and correctness of the extracted data as shown in the equation 6

$$\text{Quality\%} = \frac{TP}{TP+FP+FN} * 100 \quad \textbf{..... (6)}$$

5. EXPERIMENTAL RESULTS

The experimental results of the proposed system on the satellite images are given in the below figures. The resultant images from various steps in the proposal are shown. The gray scale image (Figure 2b) of the input file (Figure 2a) is pre-processed to get the eroded image. The erode image is the input to the Gaussian Filter which will be convoluted to remove the high frequency noises (Figure 2c).The morphological operations are performed on the Gaussian Convoluted image in order to get the road extracted binary image (Figure 2d). Using the road region pixel information from the binary image, the respective pixels are colored to get the output image road extracted (Figure 2e)

Figure 2a: Input Image

Figure 2b: Gray Scale Image

Figure 2d: Binary Road Extracted Image

Figure 2c: Gaussian Convoluted Image

Figure 2e: Road Extracted Image

Completeness, Correctness and Quality are the three performance measures used to evaluate the output. These performance measures are calculated by comparing the output image with the ground truth road as reference data. The average values for completeness, correctness and quality for the proposed system are approximately92%, 93% and 92% respectively. The output is compared with 2 other existing systems for the road extraction as shown in the Figure 3.

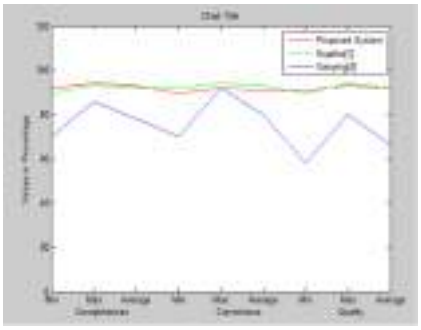


Figure 3: Comparison Plot for performance measures

6. CONCLUSION

The proposed system includes the following steps such as pre-filtering, gradient smoothening and coloring. Gaussian filtering is applied on the image in order to remove the high frequency noises. Canny edge detection algorithm is used in order to fine tune the road region edges. Morphological filling is applied on the convoluted image to obtain the road extracted binary image. The pixel values for the road region are extracted from the binary image and respective pixels are colored in the original input image. The performance measures are calculated and are compared with the existing systems to evaluate the effectiveness of the new proposed system. The values are encouraging.

REFERENCES

1.Sujatha, Selvathi, Connected component-based technique for automatic extraction of road centerline in high resolution satellite images. EURASIP Journal on Image and Video Processing. DOI 10.1186/s13640-015-0062-9 (2015)

2. C Unsalan, B Sirmacek, Road network detection using probabilistic and graph theoretical methods. IEEE Trans. Geosci. Remote Sens. 50(11), 44412434453 (2012)

3. Xiaoying, CH Davis, An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion. Inf. Fusion 6(4), 257–273 (2005)

4. Z Huang, J Zhang, L Wang, F Xu, A feature fusion method for road line extraction from remote sensing image, in Proceedings of the IEEE International Geoscience and Remote Sensing Symposium(IGARSS) (IEEE, Munich, German, 2012), pp. 52–55

5. AP Dal Poz, RB Zanin, GM d Vale, Automated extraction of road network from medium-and high-resolution images. Pattern Recog. Image Anal. 16(2), 239–248 (2006)

6. A Baumgartner, S Hinz, C Wiedemann, Efficient methods and interfaces for road tracking. Inter. Arch. Photogramm. Remote Sens. Spatial. Inf. Sci. 34(3/B), 28–31 (2002)

7. Z Miao, W Shi, H Zhang, X Wang, Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines. IEEE Geosci. Remote Sens. Lett. 10(3), 583–587 (2013)

8. X Lin, Z Liu, J Zhang, J Shen, Combining multiple algorithms for road network tracking from multiple source remotely sensed imagery: a practical system and performance evaluation. Sensors 9, 1237–1258 (2009)

9. J Senthilnath, M Rajeshwari, SN Omkar, Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method. J. Indian Soc. Remote Sens. 37, 351–361 (2009)

10. X Huang, L Zhang, Road centerline extraction from high-resolution imagery based on multiscale structural features and support vector machines. Int. J. Remote Sens. 30(8), 1977–1998 (2009)

11. H Jin, M Miska, E Chung, M Li,Y Feng, in Remote Sensing – Advanced Techniques and Platforms ed. by B resolution aerial images upon rural regions based on multi-resolution image analysis and Gabor filters. (ISBN: 978-953-51-0652-4, InTech, 2012), p. 387–416

12. T Chen, W Ji, K Zhang, A wavelet transform based method for road centerline extraction. Photogramm. Eng. Remote Sens. 70(12), 1423–1431 (2004)

13. S Movaghati, A Moghaddamjoo, A Tavakoli, Road extraction from satellite images using particle filtering and extended Kalman filtering. IEEE Trans. Geosci. Remote Sens. 48(7), 2807–2817 (2010)

14. T Geraud, JB Mouret, Fast road network extraction in satellite images using mathematical morphology and Markov random fields. EURASIP J. Appl. Signal Proc. 16, 2503–2514 (2004)

15. JB Mena, JA Malpica, An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. Pattern Recog. Lett. 26, 1201–1220 (2005)

Qualitative Improvement in Wireless Capsule Endoscopy Images using Specific Contrast Enhancement Techniques

Muddasir Khan , Ashwini S.S, M Z Kurian
Sri Siddhartha Institute of Technology, Tumkur, India.
*mudassir029@gmail.com,*ashwini_249@yahoo.co.in, mzkurianvc@yahoo.com

ABSTRACT-Wireless Capsule Endoscopy (WCE) is a non-invasive as well as an innovative solution for the detection of gastrointestinal diseases especially for small bowel diseases, such as polyps that is a main cause for gastrointestinal bleeding. The quality in these Wireless Capsule Endoscopy (WCE) images is very critical for diagnosis. The image quality is often affected with noise or degradation in acquiring and also during illumination condition. This paper presents a method to enhance the visual quality of WCE images. In this paper to improve the quality of WCE images, contrast stretching is first used to adjust the contrast of the input image. Next, Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is employed. Then, at last unsharp filtering is done to sharp the image. This yields the enhanced image which is highly improved in quality. The enhanced image highlights the details and improves the overall contrast, which is helpful for diagnosis.

Keywords— Wireless Capsule Endoscopy, WCE, CLAHE, unsharp filtering

1. INTRODUCTION

Digital images play a vital role in so many applications such as satellite television as well as in areas of research and technology such as geographical information systems etc. [1]. When an image is converted from one form to another, some form of degradation occurs at processed image. Improvement in quality of images which are degraded can be done by applying one of the enhancement algorithms. These enhancement techniques are obviously application specific. Image enhancement methods emphasize specific features to improve the visual perception in an image. Image enhancement yields an output image by varying the pixel's intensity of the input image [2].

Enhancement is basically improving the interpretability or perception of information in the images for viewers and providing much better input for other automated image processing techniques. There exist many methods that can improve a digital image without spoiling it [3]. It will be a developing trend to integrate the advantages of many algorithms to practical application to image enhancements [4].

Medical imaging is concerned with the development of the imaging instruments that aid to identify different aspects of the tissue and organs based on different properties and reveal new properties of the tissue and internal structure [5]. Medical image processing is a field of science and is gaining wide acceptance in healthcare industry because of its technological advances and software breakthroughs. It plays a critical role in disease diagnosis and in improved patient care [5].

These operations are performed in order to alter brightness, contrast or the distribution of the grey levels in image. The pixel intensities of the output image are modified according to the transformation function applied on the input values. Image enhancement is applied in every field where images are ought to be understood and analyzed. For example, medical image analysis, analysis of images from satellites etc [6].

This paper is organized as follows; Section 2 gives a brief review of previous related work and WCE imaging. Section 3 describes the proposed design methodology. Performance evaluation is presented in Section 4. Results are presented in Section 5. Conclusion is presented in Section 6.

2. RELATED WORK AND BACKGROUND

In this section, previous related work and the WCE device are explained.

2.1. Related work and contributions

Manvi, in her paper used histogram equalization for contrast enhancement in the image. Contrast enhancement methods vary the image through some pixel mapping such that the histogram of the output image is more spread than that of the original input image. The purpose of histogram equalization is to produce a uniform histogram. This operation is carried out by effectively spreading out the most frequent intensity values. Histogram equalization technique is more useful when the images with foregrounds and backgrounds that are both bright or both dark [13].

Indira et al. presented an algorithm for medical image enhancement in two stages. The first stage corrects the contrast of an image and in the second stage wavelet fusion has applied for enhancement of medical image. The restored images using this method are generally not satisfactory since the two steps adapted in this work over enhance the image. Therefore, this scheme may not be suitable for medical image enhancement [14].

R. Senthil kumar et al, in his paper presented an analysis of image enhancement techniques i.e., HE, AHE, CLAHE histogram modification methods to enhance Chest X-ray images and these are implemented in MATLAB. The performance in these algorithms is compared by using parameters like Peak signal to noise ratio (PSNR), Mean squared error (MSE), Signal to noise ratio (SNR), Absolute mean brightness error (AMBE) and Entropy. From these experimental results; and it was shown that CLAHE was

competent when compared to other techniques. CLAHE was at first developed for medical imaging and has confirmed to be successful to improve low-contrast images like x-ray images and portal films [15].

Sundaram et al. presented a histogram modified local contrast enhancement for mammogram images. This method offers better contrast enhancement and preserves local information in the mammogram images. In addition, this method leads in the detection of micro-classifications presence in the mammogram image. However, the method works well for mammogram images but it fails to produce satisfactory results for other medical images [16].

2.2. Wireless Capsule Endoscopy

Conventional endoscopy often presents limitation during diagnosis of small bowel diseases, because it is limited to the upper gastrointestinal (GI) tract, at the duodenum, and to lower GI tract, at terminal ileum. Therefore, prior to the invention of Wireless Capsule Endoscopy (WCE), the small intestine was the conventional endoscopy's last frontier, since it could not be internally visualized directly or in entirely by any techniques [7]. Wireless capsule endoscopy (WCE) is a new technology to diagnose the gastrointestinal tract problems with practically no invasiveness [8].

Wireless capsule endoscopy (WCE) is a disposable small capsule which helps to visualize the small bowel and to trace the abnormalities of the small bowel. This tiny capsule is capable to transmit color and high fidelity images of the entire gastrointestinal tract, and also intestines [9]. These WCE products have approved by the Food and Drug Administration (FDA) [10, 11]. This WCE device can be used to trace more diseases like obscure gastrointestinal bleeding, suspected Crohn's disease, chronic unexplained diarrhea, screening and surveillance for small bowel polyposis, etc. [12]. But WCE image quality is often affected with noise or degradation. Thus, the quality improvement in these images is needed in WCE medical applications.

3. PROPOSED METHOD

This section presents the proposed design methodology for the enhancement of WCE image. The flow sequence of the proposed design technique is shown in fig.1.

3.1. Contrast Stretching Operation

The input is the low contrast image which is to be enhanced. This image may have random pixel range. Therefore, contrast stretching operation is accomplished in order to translate the pixels in the display range of 0 to 255 [3].

3.2. Contrast Limited Adaptive Histogram Equalization

CLAHE operates on small regions in the image, called tiles, unlike histogram equalization, which operates on the entire image. Each tile's contrast is enhanced, such that the histogram of the output region approximately nearer to the histogram specified by the 'Distribution' parameter. The neighboring tiles are then combined by using bilinear interpolation for eliminating artificially induced boundaries. The contrast, mainly in homogeneous areas, can be limited for avoiding amplification of noise which might be present in the image. The value at which the histogram is clipped, the so-called clip limit, depends on the normalization of the histogram and thereby on the size of the neighborhood region. The CLAHE is a contrast-enhancement method that works significantly better than regular histogram equalization for most images [3].

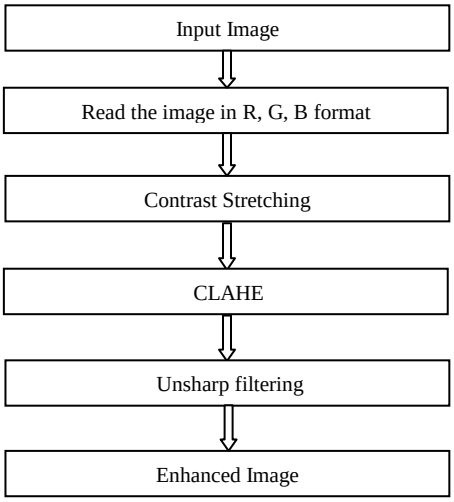


Fig. 1: Flow Sequence of the Proposed Method

3.3. Unsharp Filtering

An unsharp filter is an image sharpening operator [3]. This is one of the common techniques used for contrast enhancement. This provides better results after performing CLAHE operation. This is the final operation to get enhanced image.

4. PERFORMANCE EVALUATION

In this paper, the proposed design algorithms are implemented by using MATLAB R2013a. Performance analysis techniques require extraction of certain features that help in the identification of the object. There are many measures for quantitative analysis such as Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) etc. In this paper, PSNR, MSE, SSIM are used for performance analysis.

5. RESULTS

The results of improved WCE image compare to input image is shown in Fig 2.

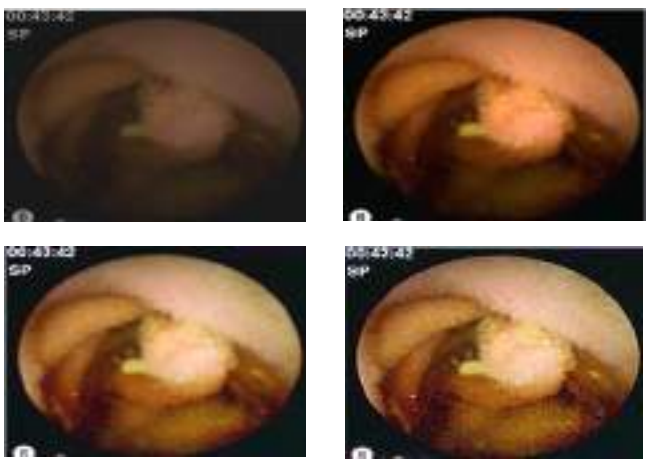


Fig 2: (a) Low Contrast Image, (b) After contrast stretching, (c) After CLAHE, (d) After unsharp filtering

Table 1: Quality metric parameter of different methods

Quality metric parameter	Proposed Method
SSIM	0.98
PSNR	32.84
MSE	33.75

6. CONCLUSION

In this paper**,** for improving the quality of WCE images, contrast stretching is first used to adjust the contrast of the input image. Further Contrast Limited Adaptive Histogram Equalization (CLAHE) is employed. Then, finally unsharp filtering is done to sharp the image. This method can highly improve the quality of the image. The proposed method can be used to enhance low contrast or poor quality medical images and that is helpful for diagnosis purposes.

REFERENCES

[1]. Dileep.MD and A.Sreenivasa Murthy, "*A Comparison between different Color Image Contrast Enhancement Algorithms*", IEEE, Proceedings of ICETECT 2011, pp 708-712.

[2]. Chen, Soong-Der and Ramli, Abd. Rahman, "*Minimum Mean Brightness Error Bi-Histogram Equalization in Contrast Enhancement*", IEEE Transactions on Consumer Electronics, Vol. 49, No. 4, 2003, pp. 1310-1319.

[3]. R.Gonzalez & R.Wood , Digital Image Processing,3rd .ed. Englewood Cliffs, NJ: Prentice Hall, 2007

[4]. Vijay A. Kotkar Sanjay.S Gharde, " *Review of various image contrast enhancement techniques*" in Image Processing, International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 7, July 2013.

[5]. K. Karthikeyan and C. Chandrasekar, "*Wavelet-based Image Enhancement Techniques for Improving Visual Quality of Ultrasonic Images*", International Journal of Computer Applications, p.p. 0975–8887, Vol. 39, No.17, Feb. 2012.

[6]. Menotti D, Najman L, Facon J and Araujo, Present a "*Multi-Histogram Equalization Methods for Contrast Enhancement and Brightness Preserving*", in IEEE Transactions on Consumer Electronics, Vol. 53, No. 3, pp.1186-1194, 2007.

[7]. Herrerías, J. M., Mascarenhas, M., 2007, Atlas of Capsule Endoscopy, Sulime Diseño deSoluciones, Sevilla.

[8]. G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "*Wireless capsule endoscopy*," Nature, vol. 405, p. 417, 2000.

[9]. D. D. Adler and C. J. Gostout, "*Wireless capsule endoscopy*," Hospital Physician, pp. 14-22, May 2003.

[10]. B. Li, and M.Q.H. Meng, "*Wireless capsule endoscopy images enhancement using contrast driven forward and backward anisotropic diffusion*", IEEE International Conference Image Processing, pp.437-440, 2007.

[11]. K. Sandrasegaran, D.D.T. Maglinte, S.G. Jennings, and M.V.Chiorean, "*Capsule endoscopy and imaging tests in the elective investigation of small bowel disease,*" Clinical Radiology, vol. 63, pp. 712-723, 2008.

[12]. F. Arguellas-Arias, A. Caunedo, J. Romero, A. Sanchez, M. Rodriguez-Tellez, F.J. Pellicer, F.Arguelles-Martin, and J.M.Herrerias, "*The value of capsule endoscopy in pediatric patients with a suspicion of Crohn's disease,*" Endoscopy, vol. 36, pp. 869–873, 2004.

[13]. Manvi, Rajdeep Singh Chauhan, Manpreet Singh, "*Image contrast enhancement using histogram equilization*," International Journal of Computing & Business Research, ISSN. 2229-6166, Proceedings of 'I-Society 2012' at GKU, Talwandi Sabo Bathinda (Punjab)

[14]. K. P Indira and R. Rani Hemamalini, "*A Method for Contrast Correction and Enhancement For Medical Images using Wavelet Fusion*", In proceedings of International Conference on Computing and Control Engineering(ICCCE), 2012.

[15]. R.Senthilkumar and Dr.M.Senthilmurugan "*triad histogram to enhance chest x-ray image*", International Journal of Advanced Research in Computer and Communication Engineering, Vol.3, Issue 11, November 2014.

[16]. M. Sundaram, K. Ramar, N. Arumugam and G. Prabin, "*Histogram Modified Local Contrast Enhancement for Mammogram Images*", International Journal of Biomedical Engineering and Technology, Vol. 9, No. 1, pp.60-71, 2012.

Prominence of Expert System and Case Study- DENDRAL

Namita Mirjankar, Shruti Ghatnatti
Karnataka, India
namitamir@gmail.com,shru.312@gmail.com

Abstract—Among many applications of Artificial Intelligence, Expert System is the one that exploits human knowledge to solve problems which ordinarily would require human insight. Expert systems are designed to carry the insight and knowledge found in the experts in a particular field and take decisions based on the accumulated knowledge of the knowledge base along with an arrangement of standards and principles of inference engine, and at the same time, justify those decisions with the help of explanation facility. Inference engine is continuously updated as new conclusions are drawn from each new certainty in the knowledge base which triggers extra guidelines, heuristics and rules in the inference engine. This paper explains the basic architecture of Expert System , its first ever success DENDRAL which became a stepping stone in the Artificial Intelligence field, as well as the difficulties faced by the Expert Systems

Keywords—Artificial Intelligence; Expert System architecture; knowledge base; inference engine; DENDRAL

I INTRODUCTION

For more than two thousand years, rationalists all over the world have been striving to comprehend and resolve two unavoidable issues of the universe: how does a human mind work, and can non-people have minds? In any case, these inquiries are still unanswered. As humans, we all are blessed with the ability to learn and comprehend, to think about different issues and to decide; but can we design machines to do all these things?

Some philosophers are open to the idea that machines will perform all the tasks a human can do. But also there are some, who openly ridicule this idea and they believe that humans are very sophisticated creatures created by nature and no machine can be equal to it.

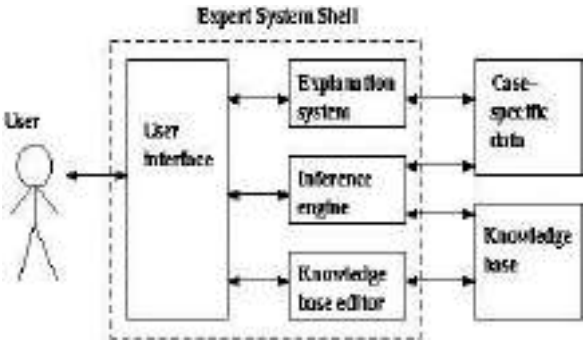
This quest to create machines which think and behave like humans has led us to the invention of Artificial Intelligence (AI). ENIAC was the first digital computer and from the time of its invention, engineers as well as other professionals have continuously tried to automate some or the other tasks in their fields. This has led us to the advancement in the field of AI that we see today. All these advancements in the field of AI and related topics can be classified into different specialized branches like robotics, expert systems, genetic algorithms, neural networks and so on.

[2] Expert system (ES) is the primary genuine business application of the innovative work completed in the AI field. As we very well know, an expert in any field is a person, who has deep knowledge, practical as well as theoretical, and also, has the proficiency in making correct judgments in taking decisions in that particular field.

An expert system is called a system, not a program, since it incorporates a few distinct parts, like, knowledge base, inference engine and user interface. All these diverse segments collaborate together in reenacting the critical thinking procedure by a recognized expert in that field. In this paper, we will study some architecture of Expert System, basic important experiments conducted, future potential and pros-cons of Expert System.

II ARCHITECTURE OF EXPERT SYSTEMS

Different architectures of Expert Systems have been proposed by different scholars and computer scientists. But the main components that remain same are: User Interface, Knowledge Base and Inference Engine.



a. Knowledge Base:

[2]The knowledge base contains the knowledge essentials for comprehension, planning and for understanding the problems arranged as schemas. Knowledge engineer is the one who creates the knowledge base by conducting a series of interviews with the experts in the specific domain. Also, he organizes the captured knowledge in a form that can be directly useful by the proposed ES. It is vital for the knowledge engineer to have the basic knowledge of terminologies and ideologies used in the proposed system.

We can categorize the knowledge captured by the knowledge engineer into three types: assembled knowledge, subjective knowledge and quantitative knowledge. The assembled knowledge is the knowledge that is captured from the advices of experts, standard experiments, previous journals, textbooks, research materials, handbooks, etc. Subjective knowledge comprises of general guidelines, estimated speculations, causal models of procedures and sound judgment. Quantitative knowledge manages systems in light of scientific speculations, numerical procedures, mathematical solutions and so forth.

Thus, knowledge base is a stockroom of the domain specific knowledge collected from different means in accordance with the information obtaining module called expert acquisition module. For the knowledge representation purpose, different rules, outlines, rationale, semantic net and so on are utilized in the knowledge base.

b. Inference Engine:

[2]Inference Engine is the most vital part of the design of any ES framework. Inference mechanisms are control systems or search strategies, which look through the knowledge base to come up with the decisions. In simple terms, we can say that, inference engine is the search module of ES.

For effective processing of ES, the inference engine works on different symbols with the help of different rules and facts; deriving the knowledge from the knowledge base. This process is recursively executed until a predetermined objective is not reached. In achieving the implementation of inference engine, many approaches are used and in them, the two most popular ones are: backward chaining and forward chaining. The main focus of backward chaining is on the final goal, whereas in the forward chaining, it is data. These two can even be combined to get a hybrid approach.

Thus, the work of the inference engine is to carry out the reasoning whereas that of the expert system is to reach to a solution

c. Explaination system:

The explanation module is the one that enables the user to ask the expert system how a particular conclusion is reached. An ES must be able to explain its actions and justify why it has concluded to a particular decision.

[4]The methods of explanations used can be classified in three categories: (a) explanation of inference on a specific input data set, (b) explanation of the knowledge base itself, and (c) explanation of the control strategy.

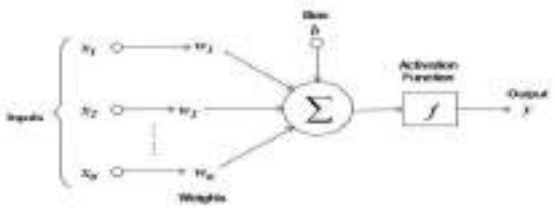
The explanation module basically answers the questions with the words: with why, how and what. Why will answer the reason, how will answer how the conclusion is reached.

d. User Interface:

It is a means through which the user communicates with the user. It makes use of different menus, graphical interface, displays etc. to make the communication easier. Obligation of the user interface is to change over the guidelines from its non-comprehendible representation to the comprehendible structure.

III ARTIFICIAL NEURAL NETWORK (ANN)

[10]A neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns.
It can be viewed in a simple way as shown in Fig.2



[12]An Artificial Neural Network consists of main four parts:

i. Inputs:

Let X = (x_1, x_2, ..., x_n), where the xi are real numbers, represent the set of inputs presented to the network model.

ii. Weights:

Each input has an associated weight that represents the strength of that particular connection. Let W = (w_1,w_2,..., w_n), with w_i real, represent the weight vector corresponding to the input vector X.

iii. Summation:

Applied to Σ, these weighted inputs along with the bias value, produce a net sum given by

$$-b + \sum_{i=1}^{n} xiwi$$

iv.Activation Function:

The value of this will be then be passed onto the activation, or *threshold* function. The threshold function might be best viewed as a single-step function, where say once the input α reaches a certain value, it will output a certain value. Next, once the threshold is activated or not, we will get an output from this neuron. We can string many neurons together and in different patterns with different layers to create a network.

Artificial Neural Network is a very vital concept for the developments in Artificial Intelligence as it resembles the human brain.

IV CASE STUDY- DENDRAL

The DENDRAL project is a good example of the rising innovation. DENDRAL was created at Stanford University to investigate chemicals. The venture was bolstered by NASA, in light of the fact that an unmanned shuttle was to be dispatched to Mars and a project was required to decide the sub-atomic structure of Martian soil, in light of the mass unearthly information gave by a mass spectrometer. The project started in 1965. Edward Feigenbaum, Bruce Buchanan and Joshua Lederberg framed a group to take care of this testing issue.

Domain	Organic chemistry- mass spectrometry
Task	To identify molecular structure of unknown compounds from mass spectra data
Input	Histogram giving mass number
Output	Description of structure of the compound
Architecture	plan-generate-test with constrained heuristic search•

Table 1

Feigenbaum had been looking for an environment in which to examine procedures of empirical induction, had arranged his reasoning toward finding such an errand situation among the exercises that researchers do. Lederberg was a geneticist. He had worked in 1965 on exobiology included the mass spectra of ammunition acids, proposed the assignment of breaking down mass spectra. Later, Buchanan also joined the team and he introduced the theory of science mixed with AI, and his interest leaned towards the scientific discoveries and their related information processes. This project was largely an experimental one. But, it became a landmark for the Expert System field.

DENDRAL stands for DENDRitic Algorithm. It is a procedure for thoroughly and non-repetitively specifying all the topologically distinct arrangements of any given set of atoms, as per the rules of chemical valence.

The central issue of diagnostic chemistry is to decide the compound structure of molecules. The general issue to

which the DENDRAL programs apply is a critical, substantive issue in the chemistry and that is structure elucidation. Structure elucidation is defined as the process of determination of the structure of a compound. The issue is critical in light of the fact that the physical as well as chemical properties of the compounds are resolved not simply by numbers of the atoms, however by their topological and geometric arrangements also. A few observational means are accessible for acquiring data about the structure of a compound. Noticeable among these is mass spectrometry and DENDRAL initially tended to the issues related just with this technique, in spite of the fact that it developed to manage the issues of structure elucidation on more broad terms.

Originally the main focus of DENDRAL was on the topologies or data of molecules, in light of the fact that the connection of its advancement was mass spectrometry. In this way to DENDRAL, diverse geometric types of the same topology were completely proportionate. In the later versions also, the basic structure of the system remained the same.

The recognizable proof of a molecule implies at any rate that its topological association is referred to; it is typically referred as to as a diagram with atoms represented as nodes and bonds represented as edges between these atoms.

At first, because specified calculation and algorithms were not yet developed for some cyclic compounds, DENDRAL was connected to aliphatic compounds only. The compounds concentrated on were amino acids, ethers, alcohols, amines etc. After developments of the algorithms, the features to DENDRAL were added consolidating the cyclic structure generator, and they worked on steroids, specifically estrogens, marine sterols, and other compounds related to it.

Roughly speaking, 100 atoms is the limit on size of molecules amiable to the DENDRAL and customary mass spectrometry investigation techniques. As of late, mass spectrometry has been effectively connected to the estimation of mass quantities of proteins with a large number of atoms. On the off chance that mass phantom investigation of pieces of proteins had been accessible, DENDRAL may have been connected to that examination (and without a doubt will be later on), utilizing super atoms to speak to individual amino acids of twenty sorts masterminded in straight successions, or to speak to DNA groupings. As it might have been, the applications were chosen to some degree for their quality in adding to the DENDRAL ideas and to a limited extent since they were of enthusiasm for their significance to contemporary chemistry.

The basic method of Heuristic DENDRAL makes use of the important concept of the generate-and-test paradigm in which a generator enumerates potential solutions, and creates all conceivable atomic structures predictable with the mass spectrum. After that, the mass spectrum is resolved or anticipated for every structure and tried against the real range spectrum. In any case, this strategy fizzled in light of the fact that a large number of conceivable structures could be created – the issue quickly got to be unmanageable notwithstanding for good estimated molecules

In addition to this, at that time, logical calculation for mapping the mass spectrum to its molecular structure was still not developed. Be that as it may, expository scientists, for example, Lederberg, could take care of this issue by utilizing their aptitudes, experience and skills. They could hugely decrease the quantity of conceivable structures by searching for surely understood examples of crests in the spectrum, and in this manner give only a couple of plausible answers for further examination. In this way, Feigenbaum's job got to be to join the aptitude of Lederberg into a software program to make it perform at a human level. Such frameworks were later named as expert systems. To comprehend and embrace Lederberg's information and work with his phrasing, Feigenbaum needed to learn essential basics in chemistry and spectral analysis. In any case, it got to be evident that Feigenbaum utilized basic rules of science as well as his own heuristics, or dependable guidelines, his own experience, and even some guessing on his part. Before long Feigenbaum recognized one of the significant troubles in the task, which he called the 'knowledge acquisition bottleneck'. He understood how difficult it is to transform information collected from human experts to apply to the computers. For this purpose, Lederberg even expected to study fundamentals in processing. In this way, Feigenbaum, Buchanan and Lederberg worked as a team and created DENDRAL, the first fruitful ES.

The main three parts of DENDRAL are: generator, planning programs and testing-ranking programs.

e. The Generator:

[7]The generator can be said as the heart of the program. The generator in the Heuristic DENDRAL is based on the DENDRAL algorithm developed by Lederberg. This algorithm determines a methodical list of molecular structures. It regards molecules as planar graphs and produces progressively bigger diagram structures until every single synthetic atom are incorporated into the graphs in every single conceivable course of arrangement. Since diagrams with cycles displayed uncommon issues,' introductory work was constrained to chemical structures without rings.

f. The planning programs:

[7]The DENDRAL Planner utilizes a lot of information of mass spectrometry to induce requirements. In the generator's lists, planning data of good and bad basic structure is put. Arranging has been restricted altogether to mass spectrometry, however the same procedures can be utilized with other information sources also.

The planning programs in DENDRAL take into consideration helpful (man-machine) critical thinking in the translation of mass spectra. It utilizes the knowledge of mass spectrometry acquired from scientists and applies it efficiently to the spectrum of an unknown molecule. That is, utilizing the scientist's meanings of the basic arrangements of the molecules and the related applicable rules, the planning programs do the accounting of fragment peaks with sections and the combinatorics of discovering predictable methods for setting substituent around them.

The output of this planning program is a structure description lists with as much detail filled in as the information and characterized fragmentations will permit.

g. The testing and ranking programs:

[7]These programs utilize a large amount of information of mass spectrometry to make testable forecasts from each conceivable applicant molecule. The predicted information is contrasted with the information from the unknown compound. These programs utilize a genuinely basic theory of mass spectrometry to predict commonly expected fragmentations for every applicant structure. Expectations which stray incredibly from the observed range are viewed as at first sight proof of error; the comparing structures are removed from the lists. Then they use more subtle rules of mass spectrometry for ranking the remaining structures.

Thus, we can summarize the Organization of the heuristic DENDRAL programs as below:

Operation	Components	Input	Output
Planning	MOLION Planning rule generator PLANNER	Mass Spectrum Planning rules Planning rules	Molecular ion constraints Constraints Superatoms GOODLIST BADLIST
Generating	Acyclic generator CONGEN GENOA STEREO	Constraints	Candidate molecular structures
Testing	PREDICTOR MSPRUNE REACT	Candidate molecular structures Mass spectrometry rules Reaction chemistry rules	Most plausible structures Structures consistent with spectrum Structures consistent with known reactions

[8]Table 2

[8]The secret to the success of DENDRAL- however not special to it- is that it attempted a very narrow, but very much characterized domain for which there was a reasonable measure of progress. The real lesson DENDRAL has for Artificial Intelligence, and for those who are intrigued by the utilization of Artificial Intelligence strategies, is that it is conceivable to select a domain of modest complexity and to decrease the issues of that domain to help the human insight. By bringing down one's sights from explaining expansive, general issues to taking care of a specific issue, by applying as much particular information to that issue as can be earned from the experts, and by systematizing and mechanizing the utilization of this information, a valuable Expert Systems can be delivered. This lesson underlies the achievement of today's Expert Systems.

V DIFFICULTIES FACED BY EXPERT SYSTEMS

Expert systems are confined to a very narrow domain. Because of this, they are not as vigorous, robust and adaptable as a user may need. Besides, ES can experience issues perceiving the boundaries of the domains. Moreover, Expert Systems have restricted explanation capabilities. They can demonstrate the rule sequence they applied to achieve an answer, however cannot relate the acquired knowledge to any more profound comprehension of the problem domain.

Verifying and validating an expert system is a very difficult job. No broad procedure has yet been created for examining their completeness and consistency. This results in difficulty in identifying incorrect, incomplete or inconsistent data.

The first generation Expert Systems had practically no capacity to gain from their experience. In addition to that, Expert Systems are assembled exclusively and can't be created quickly. It may take from five to ten man years to fabricate an Expert System framework. After so much effort, however, if the system performance and improvement depends on continuous attention from its developers then the success of the system cannot be justified.

"A good way to think about where AI fits in the entire spectrum of IT and CS is what I call the 'what-to-how' spectrum. We all know about the 'how' and AI sits at the very far end of the spectrum at the other side- the what end of the spectrum-the end of the spectrum where you would as a user tell the machine what it is you want it to do and it would have the knowledge and the reasoning power and the heuristics to employ to do it for you, so you didn't have to be a programmer, you didn't have to know any 'how'. One other things that we don't know how to do very well yet is to accumulate immense amounts of, what Doug Lenat calls- commonsense knowledge**.**

One of the great inventions of all times was writing. To write it down, to pass it on to the next generation, we move our culture to the next generation mostly by reading text, the knowledge of ordinary things, not the knowledge of specific like how you build a computer or how had the car run or something like that and we'll get that from reading text. So that's AI's number one problem today. There will be in coming up in the next ten, twenty years some really sensational computer-human interfaces in which computers can do vastly better things than they are currently doing in the service of human work. And people can do whatever residual there is that people do best and these interfaces will allow that mixture of human-computer interaction, not just where the machine is serving the person, but where the human and the computer are cooperating on a task and to profoundly greater consequences than we now think"

VI CONCLUSION

In this paper, we have identified and discussed the Expert Systems with its architecture and case study of its first success- DENDRAL. Expert System can be an extremely valuable extension of Artificial Intelligence. It can provide tremendous commercial applications in the field of medicine, agriculture, education, business accounting, legal systems, nuclear industry, and weather prediction and so on. They can be used in as simple application as offering salespersons some assistance with selling constructed homes to the complex applications like offering NASA some assistance with planning the support of a space transport in readiness for its next flight. Designing and developing an Expert System is not an easy task. It requires tremendous efforts in data acquisition, knowledge representation, application of rules etc. In spite of the fact that their utilization is far reaching, there are some professionals who are skeptic about it. As a first success, DENDRAL became the stepping stone in the field of Expert System. In the future, scientists all over the world are expecting advanced

developments in the field of Expert System in commercial as well as personal territories, which is the need of the hour.

REFERENCES

[1] Michael Negnevitsky, Artificial Intelligence, A guid to Intelligent systems 2nd ed.

[2] C. S. Krisnamoorthy and S. Rajeev, Artificial Intelligence and Expert Systems for Engineers.

[3] Expert System in Real World Applications, By Kiong Siew Wai, Abd. Latif B. Abdul Rahman, Mohd Fairuz Zaiyadi, Azwan Abd Aziz

[4] Artificial Intelligence: A Systems Approach: A Systems Approach By M. Tim Jones

[5] Figure 1: Introduction to Knowledge Engineering and Knowledge-Based Systems, *CS5201 Notes, A.P. O'Riordan, 2006*

[6] An Explanation Facility for a Grammar Writing System Loong Cheong TONG Institute of Systems Science, National University of Singapore

[7] Dendral and Meta-Dendral:Their Applications Dimension , Bruce G. Buchanan, Edward A. Feigenbaum,Computer Science Department, Stanford University, Stanford

[8] Table 2: DENDRAL: a case study of the first expert system for scientific hypothesis formation*, Robert K. Lindsay, Bruce G. Buchanan, Edward A. Feigenbaum, Joshua Lederberg

[9] INTRODUCTION TO NEURAL NETWORKS USING MATLAB 6.0, S. N. SIVANANDAM, S. N DEEPA

[10] A Brief Introduction to Neural Networks , David Kriesel

[11] An introduction to neural networks, Kevin Gurney University of Sheffield

[12] Fig. 2: http://web.eecs.utk.edu

[13] http://uhavax.hartford.edu

State of Art Technique that Accomplishes Digital Image Protection and Self Recovery

Prema C.R1, Megha S2
1Assistant Professor, 2UG Scholar
Dept. of ISE, SKIT, Bengaluru,
meghasrinivas26@gmail.com

Abstract-- Recently watermarking algorithm plays an important role in image forensics. One of the main applications is protect—ion of images against tampering. An algorithm that full-fills two purposes has to be designed that includes: 1) detection of tampered area of received image and 2) information recovery in the tampered zones. These tasks using watermarking are accomplished by state-of-the-art techniques that consist of check bits and reference bits. Check bits are used for tampering detection, while reference bits carry information about the whole image. The problem of recovering the lost reference bits can be modeled and dealt with as an erasure error. The reference bits can be protected against tampering by an appropriate design of channel code. The total watermark bit-budget is dedicated to three groups in the present proposed method: 1) source encoder output bits 2)channel code parity bits and 3)check bits. The original image is source coded and using appropriate channel encoder the output bit stream is protected in the watermark embedding phase. To retrieve the original source encoded image, erasure locations detected by the check bits help in image recovery. The image quality of both the watermarked and recovered image significantly outer-performs recent techniques as shown by the experimental results. By spending less bit-budget less bit-budget the watermarked image quality gain is achieved, where the consequence of consistent performance of designed source and channel codes result in considerably improved quality of recovered image.

Index Terms: Image watermarking, fragile watermarking, self-recovery, SPIHT, RS channel codes.

I INTRODUCTION

Digital multimedia produced are utilized in countless applications and digital imaging has been rapidly developing in last two decades. The integrity of digital images are challenged due to the popular and low cost access to image editing applications. Sophisticated techniques are required to guarantee the integrity of the image and protect it against malicious modifications. Using hash of the original image is a common approach. The image is declared unaltered if the hash output is the same as one transmitted from original image. Each image transmission reuses a secure channel for image integrity verification.

Fragile watermarking can be used for both image authentication and localization of tampered zone and recovering the image information. The integrity and localizing of tampered areas with limited robustness against image processing modifications is aimed by Inceptive fragile watermarking technique. The perfect 100% localization using watermarks robust is obtained by more recent method Watermark bits in self-recovery methods are conventionally fallen into two categories, namely check bits and reference bits. The check bits are used to localize the tampered blocks, while the reference bits are employed to restore the original image in the tampered area. Normally for the sake of content restoration, reference bits of a certain block are always embedded into another one. Nevertheless, in some of these methods content recovery may fail because both the original block and the one containing its reference bits are detected as tampered. This is called **Tampering problem**. To tackle this challenge, recent techniques spread the representation data of one block over entire image. On the other hand, there exists another problem of **watermark waste**, that is, where both original data and its reference bits are available. For instance, suggests a dual watermarking scheme where watermarked image carries two copies of content data for each block, in order to leave a chance of restoration when one copy is lost because of tampering. It should be kept in the mind that when both copies and original data survive the tampering, the watermark budget which could help the restoration of other tampered blocks is wasted. The most recent methods also deal with the watermark waste problem by offering schemes in which the content information is derived from several blocks. In our proposed algorithm, reference bits are the source coded image. This data is derived from and then scattered over the whole image to overcome both tampering and waste problems. The problem of image self-recovery is about finding an appropriate trade-off between these three parameters: the watermarked image quality, content recovery quality, and tolerable tampering rate (TTR). We approach this trade-off in our image self-recovery algorithm using these two key ideas: i) Modeling image representation and reference bit generation as a source coding problem; ii) Modeling the tampering as an erasure channel while handling it with proper channel coding. Erasure modeling of tampering has been recently

offered and exploited in and, where the authors apply fountain codes to deal with it. It should be added that when one block is marked as tampered, all its carrying reference bits are missed. We would suggest Reed-Solomon (RS) codes with large encoding blocks and over large Galva fields to solve the erasure problem. Moreover, we treat the challenge of finding some representation of the original image as a source coding problem. We apply the wavelet transform and set partitioning in hierarchical transforms (SPIHT) source encoding method to efficiently compress the original image.

Therefore, the watermark consists of three parts in our algorithm: source code bits, channel code parity bits and check bits. Source code bits which act as the reference bits are the bit stream of the SPIHT-compressed original image at a desired rate. In order to survive tampering erasure, the reference bits are channel coded to produce channel code bits. Check bits are used at the receiver to determine the erasure location for the channel erasure decoder. The output of channel decoder is source decoded to find the compressed version of the original image. This work shows that by choosing appropriate parameters for source and channel encoding, our algorithm outperforms existing methods in the same watermark payload of three bits per pixel (bpp). Nevertheless, since the watermark artifacts are significant for embedding in three LSB, we would recommend two-LSB version of our algorithm and show that its performance is still remarkable.

Organization: The paper is organized as follows. Related works is presented in section II. Our implementation modules are in section III, followed by the experimental results in section IV. We conclude at section V.

II RELATED WORK

The proliferation of digital images creates problems for managing large image databases, indexing individual images, and protecting intellectual property. This paper[1] introduces an image hash function that is a novel image indexing technique. The algorithm uses randomized signal processing strategies for a non-reversible compression of images into random binary strings, and is shown to be robust against image changes due to compression, geometric distortions, and other attacks. This algorithm brings to images a direct analog of message authentication codes (MACs) from cryptography, in which a main goal is to make hash values on a set of distinct inputs pairwise independent.

A fragile watermarking algorithm for image authentication and tamper detection is proposed [9]. A gradient image and its structure is used to achieve localization and security requirements. It provides superior localization with greater security against many attacks including vector quantization attack. In this paper [11], we introduce two techniques for self-embedding an image in itself as a means for protecting the image content. The first method is based on transforming small 8×8 blocks using a DCT, quantizing the coefficients, and carefully encoding them in the least significant bits of other, distant squares. This method provides very high quality of reconstruction but it is very fragile. The quality of the reconstructed image areas is roughly equivalent to a 50% quality JPEG compressed original. The second method uses a principle similar to differential encoding to embed a circular shift of the original image with decreased color depth into the original image. The quality of the reconstructed image gradually degrades with increasing amount of noise in the tampered image. In this study [10], an efficient self-embedding watermarking scheme for color image authentication is proposed. The scheme is designed to achieve tamper proofing and high-quality recovery. The former is used to generate authentication information for obtaining better results of tamper proofing and the latter is used to further improve the neighboring connectivity of the proofing results. The simulation results show that the proposed watermarking scheme can effectively proof the tempered region with high detection rate and can restore the tempered region with high quality.

III IMPLEMENTATION

A. Basics

The goal of our algorithm is to embed a watermark into original image to protect it against tampering. It means that the watermark must be capable of both finding the tampered areas of the received image, and recovering the content of the original image in those zones. For the purpose of image recovery, we compress the image using a source encoding algorithm, and embed the result as watermark.

However, some of compressed image information might be lost because of image tampering; hence the compressed image bit stream must be channel coded to exhibit robustness against a certain level of tampering. In order to detect tampered blocks at the receiver, some check bits are generated from those parts of image which remain unchanged during watermark embedding procedure. These check bits are inserted as a part of total watermark.

Having tampered blocks known using the check bits, tampering can be modeled as an erasure error. Therefore, compressed bit stream is channel coded using a code capable of resistance against certain level of erasure. At the receiver, the check bits locate tampered blocks. The list of tampered blocks identifies erasure locations and helps the channel erasure decoder to find the compressed image bit stream despite the occurring erasure. Then source encoded image would be decoded and the estimation of the original image is recovered.

B. Watermark Embedding

Consider the original image I represented as 8-bit gray-scale pixel values.

Fig (1). The block diagram of the proposed watermark embedding using two LSB.

These eight bits are divided into four parts: The most significant bits (MSB) that will not change at the watermark embedding phase, check bits, source code bits, and channel code parity bits. MSB bits of each pixel remain unchanged during watermark embedding and will be used later for hash generation and image reconstruction. The remaining bits are used for the purpose of watermark embedding. Block diagram of watermark embedding for 2-LSB algorithm is shown in Fig(1).

C .Tampering Detection and Image Recovery

The received image which is probably tampered is decomposed into blocks of size B × B. For each block, position bits are found, derived from shared secret key. The XOR of calculated hash bits and extracted check bits is recorded for each block. For unaltered blocks, this bit stream equals the random key used in the embedding phase.

Therefore, comparing these results and spotting the different ones leads to locating the tampered blocks. The channel decoder at the receiver side is Reed-Solomon (RS) erasure decoder. Channel code bits undergo proper inverse permutation. Then, they are delivered as input to RS erasure decoder along with the erasure locations calculated from the list of tampered blocks.

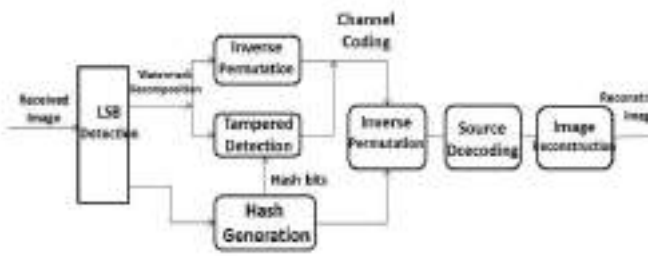


Fig (2). The block diagram of our tampering detection and image recovery scheme using 2 LSB of each pixel.

The compressed image bit stream available at the output of the decoder is passed through the source decoder after undergoing proper inverse permutation. The reconstructed image is made by replacing the tampered blocks by their corresponding blocks at the output of the source decoder. Obviously, the content of the received image in preserved blocks will not be replaced with the corresponding information derived from the restored image. An example of image recovery for 2-LSB algorithm is given in Fig (2).

IV. EXPERIMENTAL RESULTS

8-bit gray scale Cameraman image of size 512 × 512 is watermarked using our proposed method explained in Section VII. The original Cameraman image is shown in Fig. 3(a). Fig. 3(b) shows the watermarked image generated by 2-LSB version of our algorithm. As mentioned, the PSNR of watermarked image generated by 2-LSB version of our algorithm equals 44.15 dB, which is far beyond the HVS threshold of noticeable distortion. State-of-the-art tampering protection algorithms usually use three least significant bits for watermark insertion. This embedding approach degrades the PSNR of watermarked image down to 37.9 dB, which is not suitable for smooth areas. This fact is shown in Fig. 3(c), where the same image is watermarked using Zhang's method which replaces three LSB with tampering protection data. Comparing three images in Fig. 4, it is clear that Zhang's method has imposed noticeable distortion to the original image, while our watermarked image preserves the quality of the original image.

Therefore, the proposed method outperforms the state-of-the-art techniques from transparency point of view. Note that the values derived for PSNR of watermarked image (37.9 dB and 44.15 dB for those algorithms using two and three LSB for data embedding) are constant and independent of the chosen host image, in spite of the reconstruction PSNR which varies depending on the selected cover image.

Fig (3) (a) Original 8-bit gray scale Cameraman image. (b) Watermarked image generated by 2-LSB watermark artifacts are noticeable in smooth area of sky in this image.

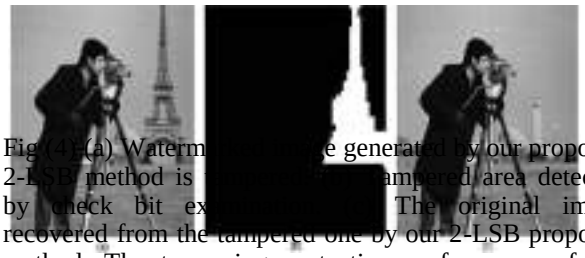


Fig (4) (a) Watermarked image generated by our proposed 2-LSB method is tampered. (b) Tampered area detected by check bit ex[illegible] The original image recovered from the tampered one by our 2-LSB proposed method. The tampering protection performance of our algorithm is also investigated in practice. Both ―low-rate‖ and ―high-rate‖ tampering scenarios are applied to Cameraman image. Fig. 4 depicts the result of low-rate tampering protection. In this case, 2-LSB version of our algorithm has been applied. Fig. 4(b) demonstrates the detected tampered areas of image after check bit

examination.

The original image is recovered exploiting the channel coded data located in preserved blocks and the list of tampered blocks as shown in Fig. 4(c). In order to protect the image against high-rate tampering, we need to spend more bit-budget for watermark embedding. Since three LSB are used for watermark data, the watermarked image looks similar to Fig. 3(c). Tampered blocks are recognized and their information is perfectly recovered as illustrated in Figs. 4(b) and 4(c).

As the next step, we compare the general performance of our algorithm with two of the most recent works presented so far. Both of them exploit three LSB for watermark embedding. In this way, regardless of the content of the watermarked image, the maximum PSNR of recovered area is limited to 40.7 dB, as is calculated. The second mode is when the tampering rate exceeds the TTR. In this scenario, the channel code breaks down, source encoder data is not retrievable and the image tampered area will not be recovered.

The results in Fig. 6 confirm the TTR calculated for 2-LSB and 3-LSB versions of the proposed method. In order to have a fair TTR comparison, we must compare the other techniques with 3-LSB version of our algorithm. It is inferred from Fig. 6 that the TTR of our proposed method is higher than that of Koru's method. Fig. 6 also confirms that our proposed algorithm dramatically outperforms Zhang's method.

The consistent performance of our proposed algorithm compared to the decaying one of Zhang's method shows significant gain in image recovery which exceeds 14 dB for high-rate image tampering. Although the notable recovery gain of our 3-LSB algorithm is attractive, since 3-LSB algorithm inherently imposes significant distortion on original images, we recommend our 2-LSB algorithm for practical applications.

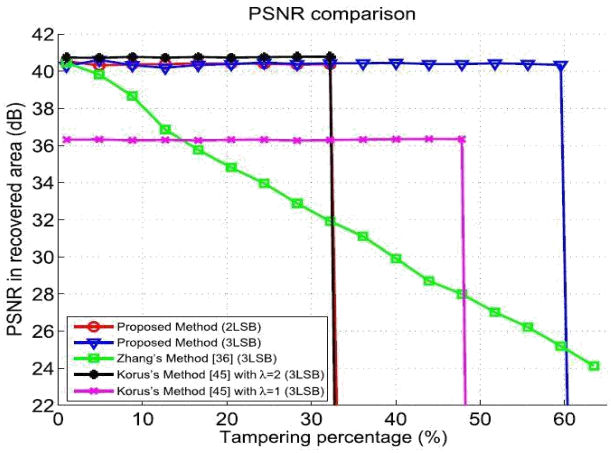


Fig. 6. Simulation results for different methods, expressed as the recovered PSNR in tampered area in terms of tampering rate.

V CONCLUSION

A watermarking scheme to protect images against tampering was introduced in this paper. The watermark bit-budget consists of three parts, check bits, source encoder output bits, and channel encoder parity bits. SPIHT compression algorithm is used for the original image source coding. The source encoder output bit stream is channel coded using RS code of a required rate and over appropriate field. The RS codes over large Galva fields are wise choices since image tampering affects a burst of bits. Check bits support the receiver in locating the tampered blocks. Therefore, the receiver knows the exact location of erroneous bits. In this way tampering is modeled as an erasure error. Thus, at the receiver we need an RS channel erasure decoder for image recovery. The lengths of the channel encoder input and output blocks are also taken as long as possible to achieve the best performance. It is shown that our watermarking scheme which replaces only two LSB of an image, efficiently recovers the tampering up to 33% without leaving any noticeable distortion. However, if we implement our algorithm using 3 LSB, it totally outperforms the state-of-the-art methods using the same three LSB for watermarking. It should be noted that albeit the proposed scheme is just implemented for two certain sets of parameters, it can be flexibly adapted to different applications with different purposes, thanks to adaptive rate adjustment capability of applied source and channel codes.

VI FUTURE ENHANCEMENT

A total of 33% tampering is recovered after the whole process is undergone so in future a whole of 100% or at the maximum of 80% tampering can be done without any distortion done to the image quality.

VII REFERENCES

[1]A. Swaminathan, Y. Mao, and M. Wu, ―Robust and secure image hashing,‖ IEEE Trans. Inf Forensics Security, vol. 1, no. 2, pp. 215–230, Jun. 2006.

[2]S. Roy and Q. Sun, ―Robust hash for detecting and localizing image tampering,‖ in Proc. IEEE Int. Conf. Image Process.(ICIP), vol. 6. Sep./Oct. 2007, pp. VI-117–VI-120.

[3]M. Wu and B. Liu, ―Watermarking for image authentication,‖ in Proc. Int. Conf. Image Process. (ICIP), vol. 2. 1998, pp. 437–441.

[4]J. Fridrich,―Image watermarking for tamper detection,‖ in Proc. Int. Conf. Image Process (ICIP), vol.

2. Oct. 1998, pp. 404–408.
[5]D. Kundur and D. Hatzinakos, ―Digital watermarking for telltale tamper proofing and authentication,‖ Proc. IEEE, vol. 87, no. 7, pp. 1167–
1180, Jul. 1999.
[6]C.-S. Lu, S.-K.Huang, C.-J.Sze, and H.-Y. M. Liao, ―Cocktail watermarking for digital image protection,‖ IEEE Trans. Multimedia, vol. 2, no. 4, pp. 209–224, Dec. 2000.
[7]P. W. Wong and N. Memon, ―Secret and public key image watermarking schemes for image
authentication and ownership verification,‖ IEEE Trans. Image Process., vol. 10, no. 10, pp. 1593– 1601, Oct. 2001.

[8]M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, ―Hierarchical watermarking for secure image authentication with localization,‖ IEEE Trans. Image Process., vol. 11, no. 6, pp. 585– 595, Jun. 2002.

[9] S. Suthaharan, ―Fragile image watermarking using a gradient image for improved localization and security,‖ Pattern Recognit. Lett., vol. 25, no. 16, pp. 1893–1903, 2004.
[10] J. Fridrich and M. Goljan, ―Images with self-correcting capabilities,‖ in Proc. Int. Conf. Image

Super Resolution Reconstruction Based on Different Techniques of Registration and Interpolation

Kamalavva Totagi, Chidananda Murthy M V, M Z Kurian
Department of Electronics and Communication, SSIT, Tumkur, Karnataka, India.
kamalatotagi33@gmail.com, chidussit@yahoo.com, mzkurianvc@yahoo.com

ABSTRACT- Resolution enhancement can effectively achieved by process of Super Resolution(SR) techniques with advantages that still Low Resolution(LR) systems can be utilized and less costly. The main object of Super Resolution is to get high resolution, high quality image from Low Resolution images. The process of imagery, the factors including the motion between earth and the platform, atmosphere disturbance, out of focus, non-ideal sampling and so on, all can make the images noisy, blurred and degraded. Super resolution technology is the signal processing based method which can detect and remove the blur and noises caused by the imaging system as well as recover information.Super resolution imaging processes one or more low resolution images acquired from the same scene to produce a single higher resolution image with more information. Recently, it has been one of the most active research areas to get high-resolution image from a low-resolution image, and for the communication purpose it is necessary to compress the information. To achieve SR, first LR images should align properly and fused to get non redundant information. Image resolution can be enhanced by interpolation. In this paper, the various registrations, fusion and interpolation algorithms are designed and implemented which can be used to compress the information which is very helpful for the communication purpose. The performance analysis can be done by PSNR and MSE. The system is implemented using Matlab GUI.

Keywords - Super resolution, Registration, RANSAC, Fusion, and Interpolation.

I. INTRODUCTION

Super resolution is a process which construct higher resolution image using one or more low resolution images of the same scene. High-quality images and videos capturing and processing is critical in many applications such as medical imaging, astronomy, surveillance, remote sensing, and so on. Traditional high-resolution (HR) imaging systems require high-cost and bulky optical elements whose physical sizes dictate the light-gathering capability and resolving power of the imaging system. In contrast, the computational imaging system is one,which combines the power of digital processing with data gathered from optical elements to generate HR images. Artifacts such as disturbance, aliasing, blurring, and noise may be affect the spatial resolution of an imaging system.

Super resolution is the process of combining the one or more low resolution images to obtain a high resolution image. The basic idea behind SR is to combine the non-redundant information contained in multiple low-resolution (LR) frames to generate a high-resolution (HR) image. A closely related technique with SR is the single image interpolation approach, which can be also used to upscale the LR image. The resolution of a digital image can be classified in many different ways such as, pixel resolution, spatial resolution, spectral resolution, temporal resolution, radiometric resolution etc. As there is no additional information provided, the quality of the single image interpolation is very much limited due to the ill-posed nature of the problem, and the lost frequency components cannot be recovered. In the SR setting, however, multiple LR observations are available for reconstruction, making the problem better constrained. The non-redundant information contained in these LR mages is typically introduced by sub pixel shifts between them. These sub pixel shifts may occur due to uncontrolled motions between the imaging system and scene, e.g., movement of objects, or due to controlled motions, e.g., the satellite imaging system orbits the earth with predefined speed and path.

The basic problem is to obtain an HR image from multiple LR images. The basic assumption for increasing the spatial resolution in SR techniques is the availability of multiple LR images captured from the same scene. In SR, the LR images represent different "looks" at the same scene. In that LR images are sub-sampled as well as shifted with sub-pixel precision. If the LR images are shifted by integer units, then each mage contains the same information, and thus there is no new information that can be used to reconstruct an HR image. If the LR images have different sub-pixel shifts from each other and if aliasing is present, and then ach image cannot be obtained from the others. In this case, the new information contained in each LR image can be exploited to obtain an HR image.

The major advantage of the super resolution approach is that it may cost less and the existing LR imaging systems can be still utilized. The SR image reconstruction is proved to be useful in many practical cases where multiple frames of the same scene can be obtained, including medical imaging, satellite imaging, and video applications. Synthetic zooming of region of interest (ROI) is another important application in surveillance, forensic, scientific, medical, and satellite imaging.

II. RELATED WORKS

Esmaeil Faramarzi, Dinesh Rajanand Marc P. Christensen.have proposed a unified blind method for multi-image super-resolution (MISR or SR), single-image blur deconvolution (SIBD), and multi-image blur deconvolution (MIBD) of low-resolution (LR) images degraded by linear space-invariant (LSI) blur, aliasing, and additive white Gaussian noise (AWGN). The proposed approach is based on alternating minimization (AM) of a new cost function with respect to the unknown high-resolution (HR) image and blurs, which improves the quality of blur. Blur deconvolution (BD) and super-resolution (SR) are two groups of techniques to increase the apparent resolution of the imaging system[1].

P B Chopade and P M Patil have proposed basic algorithms and their classification based methodology used to implement it. Due to its vast scope of applications researchers are developing a novel super-resolution algorithm for a specific intention based on single and multi-frame image resolution. In this survey, the basic concepts of the algorithms are explained and thentheir performance analyses through which each of these methods has developed are mentioned in detail[2].

Sonali Shejwal1 and Prof. A. M. Deshpande have proposed classification of SR various algorithms. Amongst which edge adaptive algorithms are particularly used to improve the accuracy of the interpolation characterizing the edge features in a larger region. This paper introduces a recent algorithm for image iterative curvature based interpolation (ICBI), and gives comparison with bicubic interpolation and the other interpolation algorithm. Comparative analysis of test images are performed on the basis of PSNR and RMSE metrics show effectiveness of edge based techniques[3].

A. Geetha Devi, T. Madhu and K. Lal Kishorehave proposed the various fusion algorithms such as averaging method, Principle component analysis (PCA) and wavelet based Fusion, scale Invariant-wavelet Transform, Laplacian pyramid, Filter Subtract decimate(FSD)pyramid[4].

Pandya Hardeep, Prashant B. Swadas and Mahasweta Joshihave presented limits on super resolution. In second part how to achieve SR and its advantages. In the first part they analyzed that super resolution becomes much more difficult as the magnification factor increases. From the analytical results of this paper which shows that the reconstruction constraints provide less and less useful information as the magnification factor increases. It is assumed that the images are noisy and down sampled [5],[7].

Min-Chun Yang and Yu-Chiang Frank Wang have proposed Learning-based approaches for image super-resolution(SR). In this paper, they present a novel self-learning approach for SR and advance support vector regression(SVR) with image sparse representation, which offers excellent generalization in modeling the relationship between images and their associated SR versions [6].

Michael Angelo Kandavalli and Raghavahave presented four main classes of methods to estimate the pixel values in HR grids, and interpolation-based approaches. In this paper, they are introduces interpolation-based approaches since both interpolation and filtering can be expressed in the form of a weighted sum. Frequency-domain approaches make explicit use of the aliasing relation between continuous Fourier transform and discrete Fourier transform [8].

Bahadir K. Gunturkand Murat Gevrekci haveproposed a Bayesian super-resolution algorithm based on an imaging model that includes camera response function, exposure time, sensor noise, and quantization error in addition to spatial blurring and sampling [9].

Niyanta Panchal, Bhailal Limbasiya, Ankit Prajapati have proposed review of different image registration methods and compare all the methods. Then next using various Super resolution methods which is generate high-resolution(HR) image from one or more low resolution images and lastly different image quality metrics reviewed as measure the original image and reconstructed image[10].

III. PROPOSED SYSTEM

The block diagram of the proposed system shown in fig: 3.1 give the high resolution image from low resolution image.

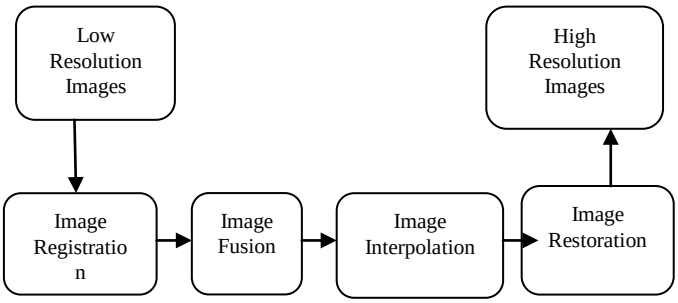


Fig 3.1: Block diagram of the system

Registration is the process of bringing all the shifted versions of low resolution images into a single plane with respect to a reference image. Feature based registration comprises of feature detection, feature matching, optimum transformation and up-sampling and provides better results in many applications. Image fusion is the process of integrating the information contained in all the low resolution observations into a single image. The resolution of the image is improved by preserving the finer details of the fused LR images during interpolation.

Different restoration technique has been employed for the reduction of noise, blur restore the images with high resolution. Several Super Resolution reconstruction algorithms are implemented. This technique can be efficiently implemented in critical applications like a medical imaging, facial recognition, bio-metrics and remote sensing to extract

the finer details of the image.The detailed description of the each module is explained below:

3.1 Noise Module:

A noise is introduced in the transmission medium due to a noisy channel, errors during the measurement process and during quantization of the data for digital storage. Each element in the imaging chain such as lenses, film, digitizer, etc. contributes to the degradation. Therefore, Noise is any undesired information that contaminates an image. Hence, the received image needs processing before it can be used in applications. A different noise model includes additive, multiplicative and impulse types of noise. They include Gaussian noise, salt and pepper noise, speckle noise and Poisson noise. A quantitative measure of comparison is provided by the peak signal to noise ratio of the image.

3.2 Image Registration:

Image Registration is the process of aligning two different images of the same scene acquired at different times, different angles, and/or different sensors. It plays an important role in remote sensing and applied in wide variety of tasks such as image fusion, image overlay and change detection using different images of the same region. Conventional image registration techniques involving manual selection of control points (CPs) and that are used to estimate the geometric transformation model and that establishes a mapping between slave image and the master. Also the manual method needs an expert with a special skill to select the individual Control Points precisely for estimating the transformation model which is a laborious activity.

An automatic image registration is a technique which can solve the pitfalls of conventional methods. The automatic image registration requires an elaborate software framework and is a very challenging task itself especially when homogeneous features such as cloud, snow features are present in the images.Image registration in automatic mode requires sequential and iterative execution of different phases forgenerating quality registered data products. The main phase or the steps in image registration are as follows:

(a) Feature Detection:The method which computes the abstractions of image information. These objects are termed as the control points which are considered to be starting point or main primitive for image registration.

(b) Feature Matching:The step which maps and establishes the correspondences between features detected in the slave image and those detected in the master image.

(c) Model Estimate:This phase estimate the transform using master and the slave coordinates by employing the models. In this step the parameters of the mapping functions and aligning the sensed image with the reference image, are estimated.

(d) Image Resampling and Transformation**:**Resampling is a process that involves the extraction and an interpolation of gray levels from pixel locations in the original distorted image and their relocation to the approximate matrix coordinate location in the corrected image. The slave image is transformed by means of the mapping function.

3.3Image Fusion:

In the field of Image processing, image fusion has received a significant attention for remote sensing, medical imaging, machine vision and the military applications. A hierarchical idea of image fusion has been proposed for combining significant information from several images into one image. The aim of image fusion is to achieve improved situation assessment and/or more fast and accurate completion of a pre-defined task than would be possible using any of the sensors individually. Mainly image fusion requires precise techniques and also good understanding of input data.

Image fusion is the process that combines information from multiple images of the same scene. The idea behind image fusion using wavelets is to fuse the wavelet decompositions of the two original images by applying fusion methods to approximations coefficients and details coefficients. By observing the performance of all the image fusion techniques, the DWT gives efficient results. Due to its orthogonality, DWT technique has been chosen for compression and decompression for the FPGA implementation of the image fusion technique.

Thus, we obtain a wavelet orthonormal basis:

$$S^{DWT} = \{\phi_{N,j}, \varphi_{1,j}, \varphi_{2,j}, \ldots\ldots \varphi_{N,j}\}_{j \in z}$$

A discrete signal **x** can be described by these scaling function and wavelet function:

1-D Wavelet Decomposition:

$$x(k) = \sum_{j \in z} s_{(N)}(j)\phi_{N,j}(k) + \sum_{i=1}^{N}\sum_{j \in z} d_{(i)}(j)\varphi_{i,j}(k)$$

Where s and d are wavelet coefficients.

3.4 Image Interpolation:

Transferring image from one resolution to another resolution without affecting the quality of image this process is called interpolation. It is also defined as approximating continues function's value using discrete samples. In the field of image processing, image interpolation is very important role for doing zooming, enhancement of image, resizing super resolution and many more. Based on considering the image features interpolation method can be classified into adaptive and non adaptive techniques. There are many interpolation techniques available, nearest neighbor, bilinear, bicubic and spline interpolation.

The general form for an interpolation function is as follows:

$$g(x) = \sum_k c_k u(d_k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{equation 1}$$

Where $g()$ is interpolation function, $u()$ is the interpolation kernel, d_k is the distance from the point consideration, X_k and c_k are the interpolation coefficients. The c_k 's are chosen

such that $g(X_K) = f(X_k)$ for all X_k. This means that the grid point values should not change in the interpolated image.

Nearest Neighbor Interpolation:

Nearest neighbor interpolation is a simplest method. It determines the gray level value from closest pixel to specified input coordinates, and assigns that value to output coordinates. This method does not interpolate values, it just copies existing values. For 2-D, the number of grid points required to evaluate the interpolation function is four.

For nearest neighbor interpolation, the interpolation kernel for each direction is:

$$u(s) = \{1 \; |s| > 0.5$$

$$\{0 \; |s| < 0.5$$

Where s = distance between interpolated point and grid point.

Bilinear Interpolation:

Interpolated point in a bilinear interpolation is filled with four closest pixel's weighted average. Two linear interpolations are performed in horizontal direction and vertical direction. It needs to calculate four interpolate function for grid point in bilinear interpolation.

For nearest neighbor interpolation, the interpolation kernel for each direction is:

$$u(s) = \{1-|s| \; |s| < 1$$

$$\{0 \; |s| > 1$$

Where s = distance between interpolated point and grid point.

Bicubic Interpolation:

Interpolated point in a bicubic interpolation is filled with sixteen closest pixel's weighted average. From this method we get sharper image than bilinear interpolation. For bicubic interpolation, the interpolation kernel for each direction is:

$$u(s) = \{3/2|s|3-5/2|s|2+1 \; 0 <= |s| < 1$$

$$\{-1/2|s|3+5/2|s|2-4|s|+2 \; 1 <= |s| < 2$$

$$\{0 \; 2 < |s|$$

Where s = distance between interpolated point and grid point.

IV. SYSTEM IMPLEMENTATION

The proposed system is implemented in GUI with the help of guide which is shown in fig 4.1. The detailed description is explained below:

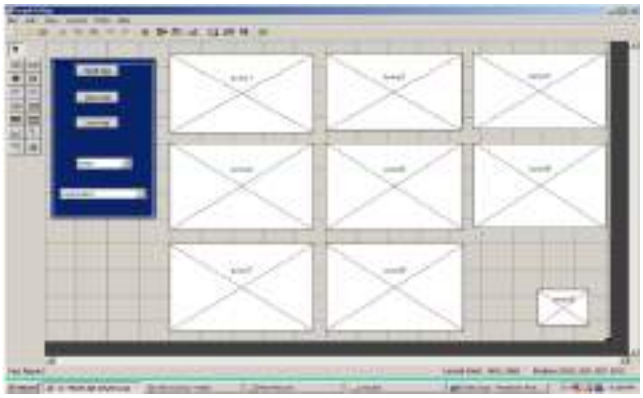


Fig 4.1: GUI guide for proposed system.

Here the image is selected either from library and then one of the various noises is added with the slider value to get the noisy image. The system also consists of popup menus to select registration algorithms.

V. RESULTS

The experimental results for registration, fusion and Interpolation algorithms with noise are shown below. Fig.5.1 shows the different noises applied to the image, which is common for both techniques.

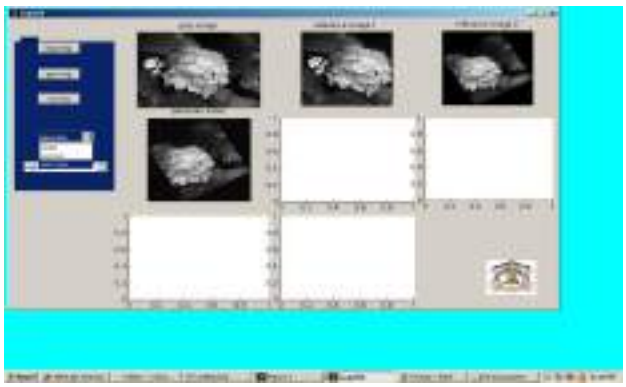


Fig 5.1: Adding different noises.

Case I: Conventional Image Registration

Fig5.2. shows the selection of control points conventionally or manually. Here input points of valid points and base points of valid points are selected.

Fig 5.2: Selection of control points.

After selection of Control Points, applying projective,affine, and non reflective similarity transformation the results is shown in Fig5.3.

Fig5.3: Conventional registration using different transformation.

Case II: Automatic Image Registration

In case of automatic image registration control points selected automatically using RANSAC method which gives inliers and outliers, they are shown in Fig5.4.

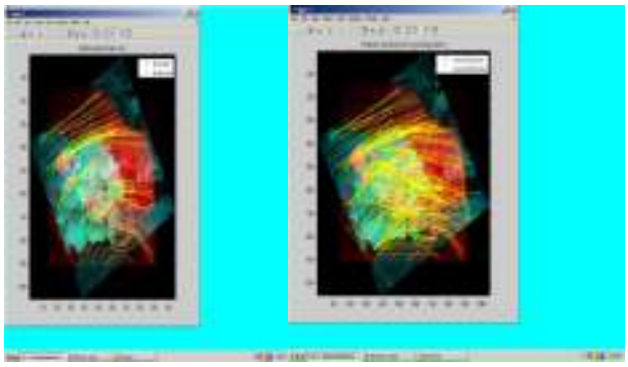


Fig5.4: Putatively matched points (inliers only& outliers).

Fig5.5. gives outputs of conventional techniques compared with automatic image registration technique using RANSAC shown below.

Fig5.5: Output conventional technique compared with automatic image registration technique using RANSAC.

Case III: Image Fusion

Fig5.6, Fig5.7 are the input images after applying fusion process we get output which is have more significant information than two input images, which is shown in below Fig5.8.

fhj

Fig 5.6: Input image 1. Fig 5.7: Input image 2.

Fig 5.8: Fused output image.

Case IV: Image Interpolation:

Different interpolation techniques shown in below Fig5.9, Fig5.10, and Fig5.11 are gives the interpolated images, with different techniques called nearest, bilinear, and bicubic to get high resolution image from low resolution images.

Fig5.9: Nearest Interpolation.

Fig5.10: Bilinear Interpolation.

Fig5.11: Bicubic Interpolation.

Table 1 : Comparison of different interpolation techniques.

Interpolation Algorithms	Computation Time	Complexity of Algorithm	Visual Quality
Nearest	Low	Low	Poor
Bilinear	Low	Low	Average
Bicubic	Average	Average	Good

Table 1 shows the comparison of nearest, bilinear and bicubic interpolation techniques based on complexity and visual quality.

VI. CONCLUSION

In this paper, the different registration methods are designed and implemented which can be used to align the images. This is a very helpful and basic step for super resolution. Image fusion gives significant information from several images into single image hence it produce necessary information to further process. Interpolation is done to enhance the resolution of low quality or resolution images based on different methods and analyzes done by comparison. In future work, an efficient and advanced super resolution algorithm will be implemented and tested.

REFERENCES

[1] Esmaeil Faramarzi, Member, IEEE, Dinesh Rajan, Senior Member, IEEE, and Marc P. Christensen, Senior Member, IEEE, *"Unified Blind Method for Multi-Image Super-Resolution and Single/Multi-Image Blur Deconvolution"*, IEEE Transactions on Image Processing, VOL. 22, NO. 6, JUNE 2013

[2] P B Chopade and P M Patil, *"Single and Multi Frame Image Super-Resolution and its Performance Analysis: A Comprehensive Survey"*, International Journal of Computer Applications (0975 – 8887) Volume 111 – No 15, February 2015

[3] Sonali Shejwal, Prof. A. M. Deshpande, *"Comparative Analysis of Edge Based Single Image Superresolution"*, International Journal of Computer Trends and Technology (IJCTT) – volume 10 Issue 5–April 2014

[4] A. Geetha Devi, T. Madhu and K. Lal Kishore, *"An Improved Super Resolution Image Reconstruction using SVD based Fusion and Blind Deconvolution techniques"*, International Journal of Signal Processing, Image Processing and Pattern Recognition ol.7, No.1 (2014), pp.283-298

[5] Pandya Hardeep, Prof. Prashant B. Swadas, Prof. Mahasweta Joshi, *"A Survey on Techniques and Challenges in Image Super Resolution Reconstruction"*, International Journal of Computer Science and Mobile Computing ISSN 2320–088X IJCSMC, Vol. 2, Issue. 4, April 2013, pg.317 – 325

[6] Min-Chun Yang and Yu-Chiang Frank Wang, Member, IEEE, *"A Self-Learning Approach to Single Image Super-Resolution "*, IEEE transactions on multimedia, vol. 15, no. 3, april 2013

[7] Simon Baker and Takeo Kanade, *"Limits on Super-Resolution and How to Break Them"*,to Appear in the IEEE Transactions on Pattern Analysis and Machine Intelligence, The Robotics InstituteCarnegie Mellon University Pittsburgh, PA 15213

[8] Michael Angelo Kandavalli * Raghava, *"An Improved Image Super-Resolution by Using Discrete and Stationary Wavelet Decomposition Via Multi-surface Fitting"*, International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 3, March 2015 ISSN: 2277 128X

[9] Bahadir K. Gunturkand Murat Gevrekci, *Student Membe, IEEE, "High-Resolution Image Reconstruction From Multiple Differently Exposed Images"*, IEEE SIGNAL PROCESSING LETTERS, VOL. 13, NO. 4, APRIL 2006 197

[10] Niyanta Panchal, Bhailal Limbasiya, Ankit Prajapati, *"Survey On Multi-Frame Image Super-Resolution"*, international journal of scientific & technology research volume 2, issue 11, november 2013 ISSN 2277-8616

A Survey on Different Methods for Brain Tumor Segmentation

NgangomPriyobata Singh,, Dr. Sunanda Dixit, Prof. B.I. Khodanpur
PGscolar, Associate Professor, Professor,
Dept. of Information Science and Engineering, DSCE, Bangalore
priyobatang@gmail.com, sunanda.bms@gmail.com, bi.khodanpur@gmail.com

ABSTRACT- Magnetic resonance (MR) images are a very useful tool to detect the tumor growth in brain. Segmentation of brain tumor from MRI images in an efficient way becomes an important role in medical image processing. One of the challenging task is to segment the tumor in an automatic manner. In this paper, we are highlighting different methods and technology used to segment the tumor. The advantages and disadvantages of several methods and technology will be discussed.

Keywords- MRI, Thresholding based segmentation, Region based method, K-Means clustering, Watershed segmentation.

I. INTRODUCTION

Image processing becomes one of the important role in the field of Medical Imaging. Magnetic resonance image (MRI) becomes an important tool for most of the researcher. It produce high quantized image giving minute details regarding delegate structure within human body. The results obtained from analysis are used to guide for the treatment.

Here a tumor may be defined as a swelling part of a body caused due to abnormal growth of tissue or cell. Studies have found that brain tumor is caused due to exposure to ionizing radiation such as radiation therapy where the machine is aim to the head and even caused due to family history. So it becomes important to detect tumor in early stage so as to give early treatment.

Brain tumor can be of two types: malignant or cancerous tumors and benign tumors. The cause of brain tumor is unknown but some of the risk factors include exposure to the industrial chemical and even as neurofibromatosis. While concern has been raised about mobile phone use, the evidence is not clear. The most common types of primary tumors in adults are: meningioma's (usually benign), and astrocytoma's such as glioblastomas. In children, the most common type is a malignant medulloblastoma. The signs and symptom of brain tumor mainly depends on size and location of the tumor. The symptom onset – in the timeline of the development of the tumor – depends in many cases on the nature of the tumor (as to it being benign or malignant). In most of the cases it is also related to the change in the nature of the neoplasm, from slow-growing, late-symptom-onset *benign* to faster-growing, early-symptom-onset *malignant*.

It has been estimated that the number of new brain cancer cases in USA, 2015 is 22,850 (12,900 males and 9,950 female) and death due to brain cancer in USA, 2015 is 15,320 (8,940 males and 6,380 female). It is found that California has the maximum number of death cases (1690 death cases) as compared to other states in USA. It is also found that the death is due to the radiation to the head (for example mobile phone), cigarette smoking and environmental toxins (for example, chemicals used in oil refineries, embalming chemicals, and rubber industry chemicals) [13].

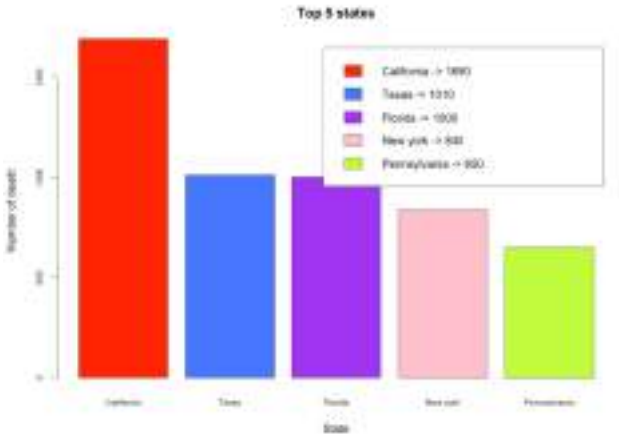


Fig 1: Graph that shows the top 5 states in which the death rate due to brain cancer is maximum in USA, 2015.

Getting segmented brain tumor from MRIs becomes very important for neurosurgeon, oncologist and radiotherapy to measure tumor responses for treatment. Segmentation can be done manually but it takes considerable time and is prone to error. So automatic detection and segmentation is highly desirable. However

automatic segmentation and detection is highly challenging task. For example, tumor can vary in size, shape and location.

II. LITERATURE OVERVIEW

***C. Senthil Singh* et al.** *[1]* have proposed a computer aided system for brain MR image segmentation for detection of tumor location using K-means clustering algorithm followed by morphological filtering. They found that this method having more advantage as it avoids the misclustered regions that can inevitably be formed after segmentation of the brain MRI image for detection of tumor location. Using these method, they were able to segment tumor from different brain MRI images. The main disadvantage of the above said method is that it produce different result for different number of clusters.

***Riries Rulaningtyas* et al.** *[2]* proposed on edge detection for brain tumor using histogram equalization and edge detection process. Three methods are suggested for edge detection. They are Robert, Prewitt, and Sobel. From these three methods of edge detection, Robert, Prewitt, and Sobel, Sobel method is more suitable for edge detection of brain tumor because it has a little mean and standard deviation value. Moreover it also gives good performance image, with edge line between brain tissues and tumor tissues are sharper than other three methods edge detection.

***Dr J. Jayakumari* et al.** *[3]* have developed a brain tumor segmentation method using Watershed Segmentation and validated using MRI Data. This method can segment a tumor provided that the desired parameters are set properly. The results show that Watershed Segmentation is the best method to segment a tumor in MATLAB environment, provided the parameters are set properly.

***J.Selvakumar* et al.** *[10]* have used computer aided method for segmentation (detection) of brain tumor based on the combination of two algorithms i.e. K-Mean Clustering and Fuzzy C-Mean. The noise free image is given as an input to the k-means and the tumor is extracted. Then segmentation is done using Fuzzy C means for accurate shape which helps in identifying types of tumor and calculation of area. Experimental result is compared with other and found that their method gives more accurate.

***Meiyan Huang* el. al** *[4]* have proposed a novel automatic brain tumor segmentation method for MRI images. In this method, tumor segmentation is treated as a classified problem. The proposed method is LIPC (local independent projection-based classification). This method used local independent projection into the classical classification model, and a novel classification framework was derived. The proposed method was evaluated using both synthetic data and public available brain tumor image data. In both problems, their method outperformed competing methods.

***Hui Tang, Huangxiang Lu* et al.** *[5]* proposed an automatic method for whole brain gliomas detection from T2W scout scan which consist of three steps. They are normalizing the pixel intensity, pixel- wise classification using a random forest classification method and finally excluding false positives in a morphological way. Themethod is evaluated using leave-one-out method and achieve an accuracy of 83.0% for HG gliomas and 78.8% for low grade gliomas in dice similarity coefficient. The result shows the potential possibility of roughly extracting brain gliomas in a single T2w MRI.

***R.Helen* el al.** *[6]* develop a Computer Aided Diagnostics (CAD) scheme for Brain Tumour detection from Magnetic Resonance Image (MRI) using active contour models and to investigate with several approaches for improving CAD performances. For segmentation, snake models, DRLSE and FCMLSM are applied. Among the three, FCMLSM is found to be the best. For classification, Random Forest and AD Tree classifiers are adopted and give sensitivity of 97% and with minimum classification error.

Parveen et al. *[7]* proposed a technique for brain tumor classification based on the support vector machine (SVM) and fuzzy c-means. The porposed algorithm is a combination of support vector machine (SVM) and fuzzy c- means, a hybrid technique for prediction of brain tumor. The image is enhanced using contrash improvement and mid-range stretch. The skull is stripped using double thresholding and morphological opeartions. The Fuzzy C means is used for segmentation to detect the tumor. Finally Grey level run length matrix(GLRLM) is employed for feature extraction, after which SVM is used to classified the tumor. The hybrid methodology of combining support vector machine and fuzzy c-means clustering for classification gives accurate result for identifying the brain tumor.

Tao Wang et al. *[9]* proposed an automatic brain tumor detection and segmentation based on a Normalized Gaussian Bayesian classification and a new 3D Fluid Vector Flow algorithm. The algorithm that is proposed by the above said authors has two major contributions. The first contribution is that Normalized Gaussian Mixture Model is used to model a healthy brain tissue. The second contribution is that they extended 2D Fluid Vector Flow

to 3D space and then used it for tumor segmentation.

Rana Banil et al. [8] proposed a method that utilizes the intensity difference immaculately by using Frequency Emphasis in Homomorphic Filtering. Noise is removed by using Gaussian filter. By thresholding, segmentation and morphological operation, the tumor regions are extracted and mapped in edge detected image of brain.

III. METHODS FOR TUMOR SEGMENTATION

A. Thresholding based segmentation

Thresholding based image segmentation aims to partition an input image into pixels of two or more values through comparison of pixel values with the predefined threshold value T individually[12]. Failure to find the most suitable algorithm to determine the threshold value(s) T the result might be one or all of the following:

1) The segmented region might be smaller or larger than the actual
2) The edges of the segmented region might not be connected
3) Over or under-segmentation of the image (arising of pseudo edges or missing edges)

B. Region based segmentation

The main goal of segmentation is to partition an image into regions. Some segmentation methods such as "Thresholding" achieve this goal by looking for the boundaries between regions based on discontinuities in gray levels or color properties. Region-based segmentation is a technique for determining the region directly.

Basic concept of seed points: **Region growing** is a simple region-based image segmentation method. It is also classified as a pixel-b. The first step in region growing is to select a set of seed points. Seed point selection is based on some user criterion (for example, pixels in a certain grayscale range, pixels evenly spaced on a grid, etc.). The initial region begins as the exact location of these seeds. The regions are then grown from these seed points to adjacent points depending on a region membership criterion. The criterion could be, for example, pixel intensity, grayscale texture, or color based image segmentation method since it involves the selection of initial seed points. Since the regions are grown on the basis of the criterion, the image information itself is important. For example, if the criterion were a pixel intensity threshold value, knowledge of the histogram of the image would be of use, as one could use it to determine a suitable threshold value for the region membership criterion[11]. Advantages of using Region based segmentation is given below

1) Region growing methods can correctly separate the regions that have the same properties we define.
2) Region growing methods can provide the original images which have clear edges the good segmentation results.

Disadvantages are

1) The computation is time consuming, no matter the time or power.
2) Noise or variation of intensity mayresult in holes or oversegmentation.
3) The method may not distinguish the shading of the real image

C.K-Means Clustering

K-Means is a least-squares partitioning method that divide a collection of objects into K groups [10]. The algorithm is given below:

1) Define k centroid, one for each singe cluster. This centroid must be put in any location on the image. The best strategy is to put each centroid as far as possible.
2) Compute the distance of each point from each cluster by computing its distance from the corresponding cluster mean. Assign each point to the cluster it is nearest to.
3) Iterate over the above two steps till the centroid does not move.

Advantages of K-Means clustering are given below:

1) Fast: K-Means and associated groundtruth can be precomputed and stored, and assigned to new data sets
2) Simple to use: Extensive hand-labeling of images by experts to generate training data is not necessary.
3) Probabilistic framework allows to approximate the actual edge probabilities for optimal segmentation

The main disadvantages of K-Means clustering is that Detection of edges still not robust enough; a single hole in a cell border can cause spillover of the watershed algorithm.

D.Watershed Segmentation

The watershed transformation considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude intensities (GMIs) correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity minimum (LIM). Pixels draining to a common minimum form a catch basin, which represents a segment [3].

The algorithm of watershed segmentation is worked as follows:

- Suppose a hole is punched at each regional local minimum and the entire topography is flooded from below by letting the water rise through the holes at a uniform rate
- Pixels below the water level at a given time are marked as flooded.
- When we raise the water level incrementally, the flooded regions will grow in size. Eventually, the water will raise to a level where two flooded regions from separated catchment basins will merge.
- When this occurs, the algorithm constructs a one-pixel thick dam that separates the two regions.
- The flooded continues until the entire image is segmented into separate catchment basins divided by watershed ridge lines.

Advantages

1. The resulting boundaries form closed and connected regions
2. The boundaries of the resulting regions always correspond to contours which appear in the image as obvious contours of objects.
3. The union of all the regions forms the entire image region

The main disadvantage of the Watershed Transform is that for most natural images it produces excessive over segmentation.

CONCLUSION

In this paper, we have presented various image segmentation techniques for brain tumor. Different tumor segmentation methods were explored and their drawbacks and advantages are presented. As per the survey, K-Means clustering is limited to only initial number of k-cluster. Wrongly selected clusters lead to erroneous result. Also for different input cluster number, different results will be given. Therefore in future, watershed segmentation is suggested which will overcome the problem of K-Means clustering.

REFERENCES

[1] Rohini Paul Joseph , C. Senthil Singh , M.Manikandan, "Brain tumor MRI imagesegmentation and detection in image processing", *International Journal of Research inEngineering and Technology Volume: 03 Special Issue: 01 | NC-WiCOMET-2014 | Mar-2014.*

[2] Riries Rulaningtyas and Khusnul Ain, "Edge detection for brain tumor pattern recognition", *Physics department, university of airlangga Surabaya.*

[3] K. S. Angel Viji, Dr J. Jayakumari, " Automatic detection of brain tumor based on magnetic resonance image using cad system with watershed segmentation", *Proceedingsof 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN 2011).*

[4] Meiyan Huang, Wei Yang, Yao Wu, Jun Jiang, Wufan Chen, senior member, ieee, andQianjin Feng, member, ieee, "Brain tumor segmentation based on local independentprojection-based classification", *IEEE transactions on biomedical engineering, vol. 61, no. 10, october 2014.*

[5] Hui Tang, Huangxiang Lu, Weiping Liu, Xiaodong Tao, "Tumor segmentation from single contrast MRI images of human brain", *Healthcare Philips Research China, 2015IEEE.*

[6] R.Helen and N.Kamaraj, 'CAD Scheme To Detect Brain Tumour In MR Images UsingActive Contour Models And Tree Classifiers', *J Electr Eng Technol Vol. 10, No. ?: 742- ?, 2014*

[7] Parveen, Amritpal singh, "Detection of Brain Tumor in MRI Images, using Combination of Fuzzy C-Means and SVM", *2015 2nd International Conference on Signal Processingand Integrated Networks (SPIN).*

[8] Rana Banik, Md. Rabiul Hasan, Md. Saif Iftekhar, "Automatic Detection, Extraction and Mapping ofBrain Tumor from MRI Scanned Images using Frequency Emphasis Homomorphic and Cascaded Hybrid Filtering Techniques.", *2nd Int'l Conf. on Electrical Engineering and Infonnation & Communication Technology (ICEEICT) 2015.*

[9] Tao Wang, I.Chang, A.Basu, " Fully automated brain tumor segmentation using a normalized Gaussian Bayesian Classifier and 3D Fluid Vector Flow.", *17th Image Processing(ICIP), 2010 IEEE.*

[10] J. Selvakumar, A. Lakshmi, T. Arivoli, "Brain tumor segmentation and its area calculation in brainMR images using K-means clustering and Fuzzy C-menas algorithm", *International conference on Advances in Engineering, Science,and Management(ICAESM), 2012.*

[11] W. Deng, W. Xiao, H.Deng, J.Liu, "MRI brain tumor segmentation with region growing method based on the gradient and variances along and inside of the boundary curve.", *3rd International conference on Biomedical Engineering and Informatics(BMEI).*

[12] A. Jyoti, M.N. Mohanty, M. Pradeep Kumar, "Morphological based segmentation of brain image for tumor detection", *International conference on Electronics and Communication System(ICECS), 20114.*

[13] Cancer Facts & Figure 2015, American Cancer Society.

An Efficient Image De-noising using Epitome and Wiener Filter

Ranjith R D, Ms. Janaki
PG Scholar, Associate Professor ,Department of Computer Science, RRCE, Bangalore-74
ranjithgowda508@gmail.com, karur.janaki@gmail.com

ABSTRACT : Picture embodiment is a scaled down, consolidated variant of the picture. It is much littler in size contrasted with the picture yet at the same time it contains the most constituent components of the relating picture. Picture encapsulation is connected to a wide assortment of picture preparing assignments, for example, Image Segmentation, Parts-based Image Retrieval, and Image In-painting and so on. Picture De-noising is additionally one of the well known uses of picture encapsulation. This article proposes a technique to enhance embodiment based de-noising. The cutting edge picture de-noising techniques use change area preparing for better commotion evacuation. This venture presents the change area preparing alongside the exemplification based de-noising system. Premise utilizing Orthogonal Locality Preserving Projection (OLPP) are learnt from the embodiment and de-noising is performed in the OLPP area.

Keywords - Domain transformation, Epitome, Image De-noising, OLPP (Orthogonal Locality Preserving Projection), PCA (Principal Component Analysis).

I. INTRODUCTION

In this undertaking we are doing Epitome based change. Recent examination is going to get self-comparability i.e. gathering together comparable patches from the entire picture, not as a matter of course adjacent has demonstrated extraordinary change in the execution of picture de-noising exhibitions. Embodiment based methodology can likewise be incorporated into this class as the exemplification contains the most constituent components speaking to the picture. This guarantees any patch of the first picture can be mapped to a patch of the comparing encapsulation. Accordingly, we get a little pool of agent patches of the picture. Another generally utilized methodology for picture de-noising depends on the changed space procedures. An arrangement of premise vectors, for example, DCT, Fourier, Wavelet, Principal Component Analysis (PCA) and so forth is utilized to extend the information/patches in the separate space and the de-noising undertaking is performed in changed area itself. The clamor free fixes are then changed back to the first spatial space. The proposed system will gives great execution when contrasted with the other change space de-noising procedures. This can be measured by execution assessment like PSNR (Peak Signal to Noise Ratio), SSIM (Structural comparability Index Metric) and MSE (Mean Square Error).

II. LITERATURE SURVEY

*"**Image de-noising by sparse 3D transform-domain collaborative filtering**"*

In this paper proposes a novel picture De-noising technique in view of an improved scanty representation in change space. The improvement of the merely is accomplished by gathering comparable 2D picture parts (e.g. obstructs) into 3D information clusters which we call "bunches". Community oriented sifting is an extraordinary method created to manage these 3D bunches. It done by three progressive steps: 3D change of a gathering, shrinkage of the change range, and opposite 3D change. The outcome is a 3D gauge that comprises of the together separated gathered picture pieces. The separated pieces are then come back to their unique positions. Since these pieces are covering, for every pixel we acquire a wide range of evaluations which should be consolidated. Collection is a specific averaging method which is abused to exploit this repetition. A sign can't change is acquired by an extraordinarily created community oriented Wiener sifting. A calculation in view of this novel de-noising methodology and its productive execution are introduced in full detail; an expansion to shading picture De-noising is additionally created. The test results show this computationally adaptable calculation accomplishes best in class De-noising execution regarding both top sign to-commotion proportion and subjective visual quality.

"Translation-Invariant De-Noising"

In this paper, graphical showcases of conventional de-noising and an alteration utilizing cycle-turning and additionally numerical tables for quantitative correlation is utilized. Things being what they are cycle-turning gives comes about that are outwardly better, frequently significantly along these lines, and quantitatively better, as almost dividing the mean-squared mistake in a few samples. We will invest a lot of energy in a particular variation: wavelet de-noising found the middle value of overall n circle shifts. This form of cycle-turning is, normally, invariant under flow movements, thus interpretation invariant - henceforth the title of the paper. The technique can be figured quickly - in n log (n) time, regardless of appearances. For the Haar-wavelet, we will likewise demonstrate that interpretation invariant

methodologies yield a few hypothetical preferences. Notwithstanding speedier rates of meeting, there is the outwardly fulfilling truth that interpretation invariant de-noising is non-oscillatory in desire. In this paper, every single computational result are reproducible.

*"**Two-Stage Image De-noising By Principal Component Analysis with Local Pixel Grouping**"*

This paper shows an effective picture using so as to de-noising plan key segment investigation (PCA) with neighborhood pixel gathering (LPG). For a superior protection of picture nearby structures, a pixel and its closest neighbors are demonstrated as a vector variable, whose preparation tests are chosen from the nearby window by utilizing piece coordinating based LPG. Such LPG method insurances, to the point that just the specimen obstructs with comparative substance are utilized as a part of the nearby measurements count for PCA change estimation, so that the picture neighborhood elements can be all around saved after coefficient shrinkage in the PCA space to uproot the clamor. The LPG-PCA de-noising strategy is iterated once again to advance enhances the de-noising execution, and the clamor level is adaptively balanced in the second stage. Trial results on benchmark test pictures show that the LPG-PCA strategy accomplishes extremely focused de-noising execution, particularly in picture fine structure conservation, contrasted and best in class de-noising calculations.

*"**Adaptive Principal Components and Image De-noising**"*

This paper exhibits a novel way to deal with picture de-noising utilizing versatile important parts. Our suspicions are that the picture is undermined by added substance white Gaussian clamor. The new de-noising procedure performs well as far as picture visual devotion, and as far as PSNR qualities, the new method thinks about PCA change estimation, so that the picture neighborhood elements can be all around saved after coefficient shrinkage in the PCA space to uproot the clamor. The LPG-PCA de-noising strategy is iterated once again to advance enhances the de-noising execution, and the clamor level is adaptively balanced in the second stage. Trial results on benchmark test pictures show that the LPG-PCA strategy accomplishes extremely focused de-noising execution, particularly in picture fine structure conservation, contrasted and best in class de-noising calculations.

III. PROBLEM STATEMENT

In this undertaking we propose a productive calculation for picture using so as to de-noising EPITOME and wiener sifting. In the current frameworks the computational multifaceted nature and nature of the yield is one of the vital issue. A few strategies can give great quality yet more calculation and the other way around. Existing frameworks utilizes wavelet change and DCT based picture de-noising. Wavelet change and DCT set aside more opportunity for the calculation. Due to that we can't accomplish sifting as for time. This downside can be overcome by EPITOME change and wiener separating.

IV. METHODOLOGY

1. Image de-noising From Noisy Image
2. Epitome
3. Orthogonal Locality Preserving Projection (OLPP)
4. Image Blocking
5. PSNR,MSE performance

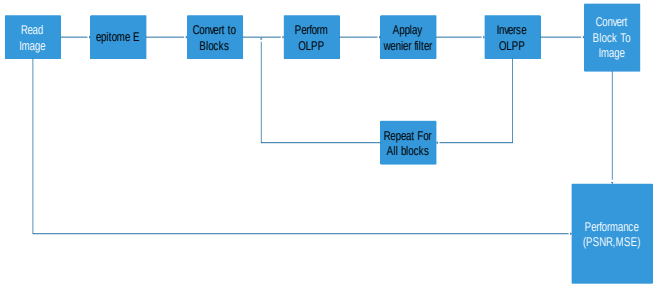


Figure 1: Block Diagram of Existing Method

The accompanying graph demonstrates the engineering of proposed technique. In this first picture is connected to EPITOME. Encapsulation is a change used to discover the properties of framework. At that point picture is changed over into little pieces to locate the neighborhood district properties. At that point OLPP is performed to extend the picture pieces. Fundamentally it will build the commotion breaking down capacity. At that point apply wiener separating to evacuate the commotion. When you connected OLPP we need to apply opposite change to get into spatial space. The same strategy is rehashed for every one of the squares. At that point at last opposite change and reverse blocking is performed to recover the picture. PSNR, SSIM and MSE is discover for the execution.

V. ALGORITHM

Input: Noisy Image (X), patch size (p), noise variance (σ).

Output: De-noised Image (X1)

1) Generate epitome E from the input noisy image X..
2) Learn global OLPP basis from the epitome E.
3) Repeat steps 4-7 for all overlapping patches of input image X, i.e. x1, x2,...,xn of size p.
4) Project the patch xi in the OLPP domain

$$y_i = a^T x_i \quad \text{.....(1)}$$

5) Apply Weiner filter update

$$\hat{y}_i = \frac{\sigma_{y_i}^2 - \sigma^2}{\sigma_{y_i}^2} y_1 \quad \text{.....(2)}$$

6) Inverse transform the de-noised patch $\hat{y}_i$

$$\hat{x}_i = a\hat{y}_i \quad \text{.....(3)}$$

7) Place the de-noised patch xi on its corresponding spatial location.

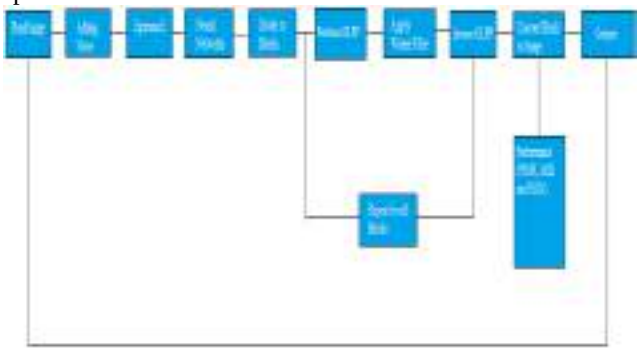


Figure 2: Block diagram of Proposed Method

VI. RESULTS AND DISCUSSIONS

Figure 3: PSNR Performance (X axis-Noise Ratio, Y axis-PSNR Value)

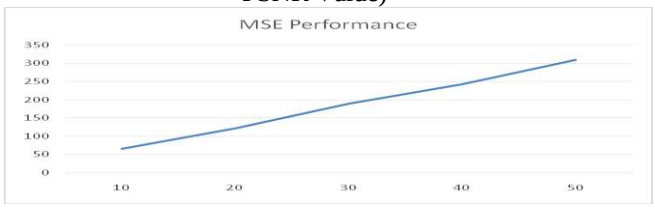


Figure 4: MSE Performance(X axis- Noise Ratio, Y axis-MSE values)

Figure 5: SSIM Performance(X axis- Noise Ratio, Y axis- SSIM values)

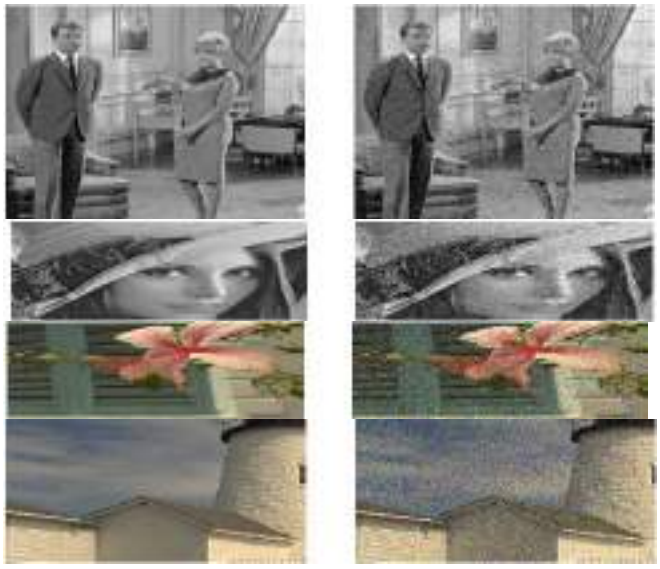


Figure 6: RHS images are Noise Images and LHS images are De-noised Images.

Image	Existing Method		
	PSNR	SSIM	MSE
Family			
Leena			
Hibiscus			
Home			

Table 1: Existing method for Greyscale images.

Image	Proposed Method		
	PSNR	SSIM	MSE
Family			
Leena			
Hibiscus			
Home			

Table 2: Proposed method for Greyscale Images.

Image	Existing Method		
	PSNR	SSIM	MSE
Family			
Leena			
Hibiscus			
Home			

Table 3: Existing method for Colour images.

Image	Proposed Method		
	PSNR	SSIM	MSE
Family			
Leena			
Hibiscus			
Home			

Table 4: Proposed method for Colour images.

VII. CONCLUSION

Picture exemplification is a small form of the picture which contains the most Constituent components of the

comparing picture. It takes a shot at the idea of collection comparable patches. While producing Utilizing the embodiment gives a picture with preparing for better clamor evacuation. We proposed a technique for utilizing OLPP to enhance embodiment based. The embodiment of the loud picture, which is much littler in size contrasted with the picture, is utilized to take in the OLPP premise and afterward the picture is restored utilizing these OLPP premise. This technique for picture de-noising is investigated diverse clamor levels. The outcomes propose a significant measure of change over the first embodiment based de-noising. Exemplification based change area.

REFERENCES

[1] R. R. Cofiman and D. L. Donoho, "Translation-Invariant de-noising." Springer-Verlag, 1995, pp. 125-150.

[2] F. Chung, "Spectral graph theory," in Regional Conf. Series in Math. no. 92, 1997.

[3] B. J. Frey and N. Jojic, :Learning the "epitome" of an image," Technical Report PSI-2002-14, University of Toronto Technical Report, Tech.Rep., 2002.

[4] D. D. Muresan and T. W. Parks, "Adaptive principal components and image denoising," in ICIP (1), 2003, pp. 101-104.

[5] X. He and P. Niyogi, "Locality preserving projections," in MIPS, 2003.

[6] N. Jojic, B. J. Frey, and A. Kannan, "Epitome analysis of appearance and shape," in Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003, pp. 34-41.

[7] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment error visibility to structural similarity," Image Processing, IEEE Transactions on, vol. 13, no.4, pp. 600-612, April 2004.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol.16, no. 8, pp. 2080-2095,2007.

[9] A. Buades, B. Coll, and J. Morel, "Nonlocal image and movie denoising, International Journal of Computer Vision, vol. 76, no. 2, pp. 123–139, 2008.

[10] L. Zhang , W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," Pattern Recognition, vol.43, no. 4, pp. 1531-1549, April 2010.

[11] X. Chu, S. Yan, L. Li, K. L. Chan, and T. S. Huang, "Spatialized epitome and its applications," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 311-318.

[12] G. Shikkenawis, S. K. Mitra, and Rajwade, "A new orthogonalization of locality preserving projection and applications," in Pattern Recognition and Machine Intelligence. Springer, 2013, pp. 277-283.

SPEECH *CONTROLLED* SMART WHEELCHAIR USING REGIONAL LANGUAGES

Chandramma R[1],Akshatha K Gowda[2], Asha U[2],Jyothi B S[2] ,Meghana B Bapat[2] ,
Department of Computer Science and Engineering,
Vivekananda Institute of Technology, Bangalore-74
rchandramma.vkit@gmail.com,akshathagowda34@gmail.com,ashanayak919@gmail.com
sudhindra.jyothi@gmail.com,meghna.anagha@gmail.com

***Abstract*:** People with physical disability or subjected to other injuries who cannot walk will use wheelchairs. In today's world, development is so enhanced that it assures to develop a smart wheel chair. This paper presents a smart wheelchair which is developed to monitor the movement of wheelchair based on speech using regional languages for physically disabled people. To monitor this wheel chair inbuilt speech functions of regional languages are used. This system allows the user to robustly interact with the wheelchair at different levels of the control and sensing.

Keywords: Smart Wheelchair, Physically disabled, Regional languages, Speech function, Microcontroller

I. INTRODUCTON

1.1 Background scenario

According to World Report on disability [1] presented by World Health Organization (WHO), there are 70 million people handicapped in the world. Nowadays due to road accident and disease like quadriplegics the number of handicapped people is increasingdrastically. Percentage of physical disabilitystands first among all disabilities. Usually, handicapped people will have to be dependent on other people for their daily activities like transport, food and movement.

1.2 Indian Statistics on Disabilities

At every moment population of India is increasing rapidly. In India 120 million people are disabled in which 41.32% are physically handicapped.

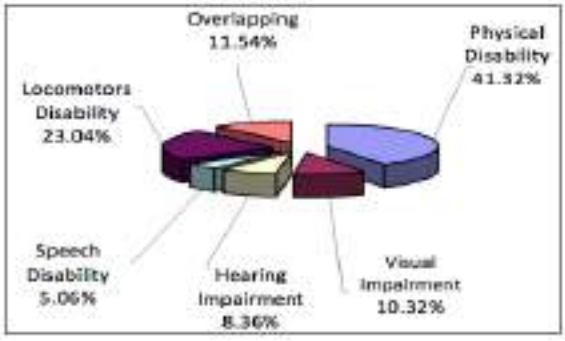


Fig 1: Statistics of Physical Disability in India

The aim of the project is to use wheelchair automatically and operate using voice through mike for moving forward,backward, left and right as well [2]. Quadriplegics and Multiple Sclerosis patients have severe disabilities and cannot drive joystick operated traditional wheelchair [5]. The limitation oftraditional wheelchair is bulkiness, flexibility and limited function.

A regional language is a language spoken in an area of sovereign state, whether it be a small area, a federal state or province, or some wider area. People who are illiterates, who does not know English will communicate through regional languages.

A wheelchair fitted with Obstacle Sensor, Motor and Mike to help driver to achieve independent mobility. By just giving commands using regional languages through mike the wheelchair can be moved in 4 directions. The obstacle sensor can help the rider control the wheelchair by taking over some of the responsibility for steering and avoiding objects until the user is able to handle the job. The approach allows the user to use human voice synchronize with the movement of wheelchair so that they can use it with comfort.

The complexity is reduced by making use of mike. So that the size of the system is very compact.

Wheelchair is integrated with voice so that handicapped people who cannot walk, who does not have hands can move chair by voice commands through mike.

Taking advantage of technological evolution in order to increase the quality of life for handicapped people and facilitate their integration into the working world [2].

II. EXISTING METHODOLOGY

(a) Masato Nishimori, Takeshi Saitoh and RyosukeKonishi(2007) Voice commands are used for mobility of the wheelchair. Voice recognition is based on reaction commands which is given through headset microphone and a laptop.

(b) K.Sudheer, T.V Janardhanarao, Ch.Shridevi M.s Madhan Mohan(2012) Voice and gesture based electronic powered wheelchair using ARM used combination of speech and gesture recognition .In this speech recognition module ,hidden markov model are used. The MEMS sensor is used and it senses the angle of hand . For Voice recognition the voice IC is used.

(c) M. Prathyusha ,K.S Roy, Mahboob Ali sheikh(2013 April) Voice and touch screen based direction and speed control of wheelchair. The speech recognition system uses programmable speech recognition circuit. The speed controller works by varying the average voltage sent to the motor.

(d) Rakhi A. Kalautri , D.K Chitre (2013) Used automatic gesture recognition system based on acceleration sensor here used is 2–axis .By calculating amount of tilt and output of tilt will decide to more in which direction.

III. APPLICATIONS OF SMART WHEELCHAIR

-Hospitals
-Sports
-Physically handicapped individuals

IV.ARCHITECTURE OF SPEECH CONTROLLED SMART WHEELCHAIR

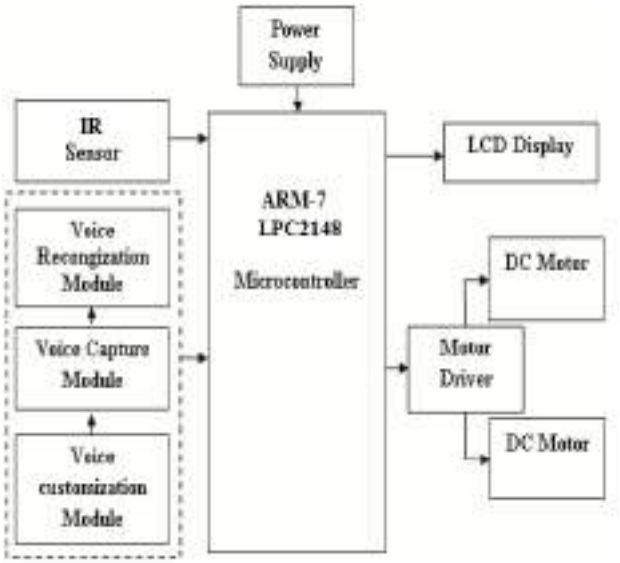


Fig 2: Block diagram of speech controlled smart wheelchair

Fig 2 presents block diagram of proposed Speech controlled Smart Wheelchair. There are 6 modules in it. They are power supply module, speech recognition module, Infrared module, LCD display module, Driver motor interface module and Microcontroller module. Microcontroller acts as a main controller. Input of 5V is provided to microcontroller through power supply module which can drive the remaining modules. The speech recognition module contains 3 sub modules in it, which are Voice recognition module, Voice capture module and Voice customization module. For training the kit to receive sound commands signals in which user gives customized commands using keypad present in the kit which can lead directions of the wheelchair such as FORWARD, BACKWARD, RIGHT, LEFT AND HOLD ON is given by Voice Customization module. The microphone present in the speech kit for capturing the speech commands is utilized by Voice Capture module. Then it will convert the received commands into binary codes according to its frequency of the speech command which are compared with predefined commands stored in the microcontroller. Comparing the binary codes converted with the one stored in the microcontroller then utilizing these command if both are alike is handled by Voice Recognition module.

To sense certain characteristics of surroundings Infrared Sensors are used. It senses either by emitting and/or detecting through Infrared radiations.In our project, it also acts as an obstacle detector.If obstacles are detected then it switches off the system immediately and the buzzer rings.

We also make use of LCD [Liquid Crystalline Display]. It is used to display the given commands.

V. RESULTS

The prototype system is implemented on small wooden chair. In this project, we make use of Regional Languages which will help people to communicate and operate the wheelchair in their own languages. Infrared sensor is interfaced with the microcontroller which detects obstacles and calculate the distance from the wheelchair and switch off the movement immediately and rings the buzzer. Under the back wheeled foot of wheelchair 2 DC motors are fixed which drive the wheelchair in different directions. The speech recognitions, driver interface, microcontroller, infrared sensor and power supply modules are implemented using Embedded C programming. To drive the DC motors in clockwise and anticlockwise directions, the microcontroller is programmed based on the requirement. The DC motor is controlled by 4 relays for themovement of the wheelchair according to the current commands received from the microcontroller. For forward and backward movement; both DC motors are rotated in clockwise and anticlockwise direction, respectively. For turning left, the left motor moves anticlockwise and for right movement, right motor moves clockwise on. Finally if an obstacle is to be detected by infrared detector, the microcontroller sends signals to both DC motors to stop operating immediately and the buzzer rings. In case of Emergency, Panic button is also included when pressed the buzzer rings. LCD module displays the command from the microcontroller.

Serial No.	COMMANDS	DIGITAL SIGNALS	ACTIONS PERFORMED
1	Mundhe	1010	Forward
2	Hindhe	0101	Backward
3	Eda	0010	Left
4	Bala	1000	Right
5	Edakke tirugu	0110	Instant Left
6	Balakke tirugu	1001	Instant Right
7	Nillisu	0000	Stop
8	Mundu	1010	Forward
9	Venaka	0101	Backward
10	Edamu	0010	Left
11	Kuda	1000	Right
12	Edam vaipu	0110	Instant Left

	tirugu		
13	Kuda vaipu tirugu	1001	Instant Right
14	Nilupu	0000	Stop
15	Front	1010	Forward
16	Back	0101	Backward
17	Left	0010	Left
18	Right	1000	Right
19	Instant Left	0110	Instant Left
20	Instant Right	1001	Instant Right
21	Stop	0000	Stop

TABLE I: List of Commands on which Speech controlled Smart Wheelchair operates

VI. CONCLUSION

By using regional languages majority of the people can operate the wheelchair comfortably. Microcontroller LPC2148 is programmed to move the wheelchair in all required directions. The speech recognition kit identifies the voice instruction provided to move the chair according to given directions or to stop. Infrared detector module detects the intruders that may appear in the path and stops the wheelchair for the response of identifying the intruder. LCD display is used to show the logic operation performed by the wheelchair. The DC motor is driven through 4 relays by the microcontroller instructions. The proposed wheel chair functions as a voice controlled load carrying robot that is very useful for aged and physically challenged people.

Instead of using voice recognition can use eye retina using optical sensor to move wheelchair in different directions. Tongue operated assistive technology is possible to access to android phone applications using Bluetooth link.

REFERENCES

[1] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] P.Sutha, S. Prabhu, S. Manikandan, S. Venkateshkumar, A. Stephen Paul. International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 12, June 2013

[4] Rajesh KannanMegalingam, Ramesh Nammily Nair, SaiManojPrakhya, *" Automated Voice based Home Navigation System for the Elderly and the Physically Challenged", Feb. 13~16, 2011 ICACT201, pp.603*

[5] 608Rakhi A. Kalantri, D.K. Chitre*Automatic Wheelchair using Gesture Recognition*, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 9, March 2013

[6]*"Android Phone Controlled Voice, Gesture and Touch Screen Operated Smart Wheelchair"* (ShraddhaUddhavKhadilkar, NarendraWagdarikar, E and TC Department, Dnyanganga College Of Engineering and Research, Pune, India,2015)

[7] J.D.Nisbet, I.R. Loudon and J.P. Odor, "The *CALL Centre smart wheelchair".* In processing of *1st International Workshop on Robotic Applications to Medical and Health Care*, 1988, Ottawa, Canada

[8] HakanNeveryd and Gunnar Bolmsjo, "*WALKY, A mobile robot system for the disabled"*. In Proceedings of the 4th International Conference on Rehabilitation Robotics (ICORR), June, 1994, Wilmington, Delaware, USA, pp. 137-141

[9] "*Speech Recognizing Powered Wheelchair for Disabled*" (D.K. Prathiba, Chidananda Murthy. M. V & M. Z. Kurian Dept. of Electronics and Communication, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India)

[10] "*Voice Recognition and Touch Screen Control Based Wheelchair for Paraplegic Persons"* (Aruna.C, DhivyaParameshwari. A, Malini. M, Gopu. G, Sri Ramakrishna Engineering College, Coimbatore, India)

[11] M SenthilSivaKumar, JaykishanMurji, Lightness D Jacob, Frank Nyange, M. Banupriya "Speech Controlled Automatic Wheelchair" Pan African International Conference on Information Science, Computing and Telecommunications, 2013.

Study and Analysis of DWT-SVD Based Digital Image Watermarking Technique for Colour Images

Rebecca. A, Usha Sakthivel
Department of Computer Science & Engineering,
RRCE, Bangalore – 560074, Karnataka, India
rebecca.engg@gmail.com, sakthivelusha@gmail.com

ABSTRACT: This paper displays a powerful and visually impaired advanced picture watermarking procedure to accomplish copyright security. Keeping in mind the end goal to shield copyright material from unlawful duplication, different advances have been produced, similar to key-based cryptographic strategy, computerized watermarking and so on. In computerized watermarking, a mark or copyright message is furtively inserted in the picture by utilizing a calculation. The calculation of computerized combining was actualized so as to watermarking both DWT and SVD methods. At first, cover picture is decayed into 4 sub-groups utilizing 2-D DWT, and later SVD is applied on every band by adjusting their solitary qualities. Subsequent to subjecting the watermarked picture to different assaults like obscuring, including commotion, pixilation, revolution, rescaling, contrast alteration, gamma redress, histogram evening out, editing, honing, lossy pressure, the initially embedded watermark picture from every one of the groups is removed and analyzed on the premise of their MSE and PSNR values. Exploratory results are given to delineate that on the off chance that we perform alteration in all frequencies, then it makes watermarked picture more impervious to an extensive variety of picture preparing assaults i.e. watermark from any of the four sub-groups can be regrouped productively.
Keywords: Digital image watermarking, DWT and SVD, Key-based cryptographic, MSE and PSNR value.

1. INTRODUCTION

Steganography and watermarking are the two techniques which can be utilized to install data straight forwardly into these substance. Not at all like standardized identifications, can't they take away from the feel of a picture. Watermarks are indistinguishable from the principle content in which they are implanted. At long last, watermarks experience the same changes as the primary substance. The execution of the watermarks can be assessed on the premise of little arrangement of properties such as heartiness, loyalty, and vagueness and so on.

Watermarking plans can be separated into two primary classes as indicated by the implanting area: spatial and change space[10]. Since high frequencies will be lost by pressure or scaling, the watermark sign is connected to the lower frequencies, or even better, connected adaptively to frequencies that contain imperative data of the first picture. In DWT-based watermarking strategies, the DWT coefficients are altered to watermark information. The alteration is generally made in high frequency sub bands to keep up better picture quality.

This paper exhibits a vigorous and visually impaired computerized picture watermarking procedure to accomplish copyright assurance. Keeping in mind the end goal to shield copyright material from unlawful duplication, different advances have been produced, similar to key-based cryptographic procedure, computerized watermarking and so forth. In computerized watermarking, a mark or copyright message is furtively installed in the picture by utilizing calculation. In our paper, we execute that calculation of computerized combining so as to watermarking both DWT And SVD systems. At first, we disintegrate the first (cover) picture into 4 sub-groups utilizing 2-D DWT, and afterward we apply the SVD on every band by changing their particular qualities. In the wake of subjecting the watermarked picture to different assaults like obscuring, including commotion, pixilation, revolution, rescaling, contrast alteration, editing, honing, lossy pressure and so forth, we remove the initially embedded watermark picture from every one of the groups and look at them on the premise of their Mean Square Error (MSE) and Peak Signal Noise Ratio (PSNR) values.

Test results are given to outline that, on the off chance that we perform adjustment in all frequencies, then it will make our watermarked picture more impervious to an extensive variety of picture handling assaults (counting regular geometric assaults), i.e. we can recoup the watermark from any of the four sub-groups proficiently.

2. LITERATURE SURVEY

Computerized watermarking innovation is an outskirts research field and it serves an essential part in data security. As indicated by the investigation of the definition and fundamental qualities of advanced watermarking innovation, the framework model of computerized watermarking is given.

The framework comprises of two modules which are watermark installing module and watermark discovery and extraction module. In perspective of the significance of advanced pictures copyright security, in light of the

investigation of the primary computerized watermarking calculations, the computerized watermarking innovation can be connected to the picture copyright insurance. The two measurement discrete cosine change is encoded on the Windows stage by utilizing Visual C++ program dialect. The analysis result demonstrates that the computerized watermark is non-detectable; the watermark data can be removed regardless of the possibility that it has been assaulted, and the normal impact can be accomplished [1].

Outlining minimal effort and rapid verification answer for computerized pictures is dependably an alluring range of examination in picture preparing. In this paper a sum up picture confirmation strategy has proposed by hybridization of shading histogram and related initial four factual minutes to accomplish the goals of minimal effort and fast. Proposed strategy can apply for both dark and shading pictures having any size and any configuration. Arrangement creates a little validation code without hardly lifting a finger means which is use to dissect the attributes of got picture from altering point of view [2].

Watermarking systems which are delicate to purposeful alterations while hearty to coincidental or accidental controls are alluded to as Semi-delicate. This paper proposes a semi-delicate watermarking procedure which installs watermark signal into the host picture keeping in mind the end goal to validate it [3].

The watermark is created as a parallel example from the element of the host picture and is inserted in the high recurrence sub band in the wavelet space. Top Signal to Noise Ratio (PSNR) and Similarity Ratio (SR) are registered to gauge picture quality. Recreation results demonstrate that this method still jelly high picture quality after the inserting prepare and is strong against a percentage of the coincidental picture handling operations while showing the fraud if the picture is intensely prepared.

The test on account of picture verification is that as a rule pictures should be subjected to non vindictive operations like pressure, so the confirmation procedures should be pressure tolerant. In this paper we propose a picture validation framework that is tolerant to JPEG lossy pressure operations. A plan for JPEG dim scale pictures is proposed in view of an information installing strategy that depends on a mystery key and a mystery mapping vector in the recurrence area. An encoded highlight vector extricated from the picture DCT coefficients is inserted needlessly and imperceptibly in the stamped picture [4].

On the recipient side, the element vector from the got picture is inferred again and contrasted against the extricated watermark with check the picture legitimacy. The proposed plan is strong against JPEG pressure up to a greatest pressure of roughly 80%, yet touchy to malignant assaults, for example, cutting and sticking.

3. DESCRIPTION AND APPROACH FOLLOWED

3.1. DWT

The wavelet space has turned into an appealing area for the watermarking of advanced pictures because of its well coordinating conduct with human visual framework (HVS) [5].

The DWT forms the picture by partitioning it into four non covering multi-determination sub-groups LL, LH, HL and HH [6]. The sub-band LL speaks to the coarse-scale DWT coefficients (the estimate) whiles other sub-band represent the fine-scale of DWT coefficients (the details).

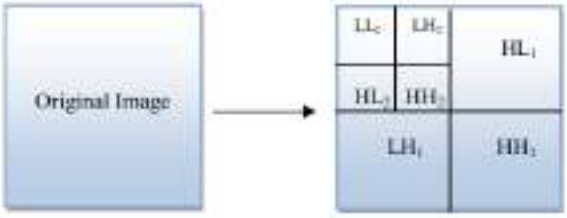


Figure 1: Two level DWT decomposition

3.2. SVD-based Watermarking

The solitary worth decay (SVD) of m×n genuine esteemed lattice A with m _ n, performs orthogonal line and section operations on An in a manner that the subsequent grid is askew and corner to corner qualities (particular qualities) are orchestrated in diminishing esteem and match with the square base of the Eigen estimations of A^TA [7]. The section of the m×m, U has commonly orthogonal unit vectors, just like the segments of the n×n, V network. U and V are orthogonal networks i.e.

$$U^TU = V^TV = VV^T = I$$

4. OBJECTIVE OF THE PROJECT & DESCRIPTION

The target of the task is to build up a watermarking strategy to shield copyright material from unlawful duplication.

In this paper have acquainted DWT-SVD procedure with implant watermark picture into the principle or spread picture, which demonstrates hearty to different sort of assaults.

DWT The wavelet space has turned into an alluring area for the watermarking of computerized pictures because of its well coordinating conduct with human visual framework (HVS)[9]. It is utilized as a part of assortment of sign handling applications, for example, video pressure; Internet interchanges pressure, object acknowledgment and numerical examination. The principle highlight of DWT is multi-scale representation of capacity.

4.1. SVD-based Watermarking

The particular quality decay (SVD) of m×n genuine esteemed framework A with m * n, performs orthogonal

line and segment operations on and in a manner that the subsequent network is inclining and corner to corner qualities (solitary qualities) are orchestrated in diminishing esteem and harmonize with the square foundation of the Eigen estimations of A^TA[11]. The segment of the m×m, U has commonly orthogonal unit vectors, just like the sections of the n×n, V framework. U and V are orthogonal frameworks[8].

4.2. Advantages

Protect the copyright material from illegal duplication.

DWT and SVD techniques are used to get accurate high quality image.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The sizes of the particular qualities for every sub-band of the colour picture are appeared in the Fig. 2 demonstrates 512×512 colour picture , the 256 × 256 dim scale visual watermark copyright, the watermarked picture, and the watermarks developed from the four sub-groups. The scaling variable i.e. _k for LL sub-band is taken to be 0.05 and 0.0005 for other three sub-groups.

Our executed plan depends on supplanting particular estimations of the HH band with the solitary estimations of watermark. The wavelet coefficients are found to have biggest worth in LL band and most minimal for HH band.

Figure 2: Cover image

Figure 3: Watermark image

Figure 4: Watermarked image

5.1. Performance Analysis

Watermark embedding in LL band is resistant to attacks including Gaussian noise, salt & pepper noise, mirroring (both vertical as well as horizontal), and JPEG compression.

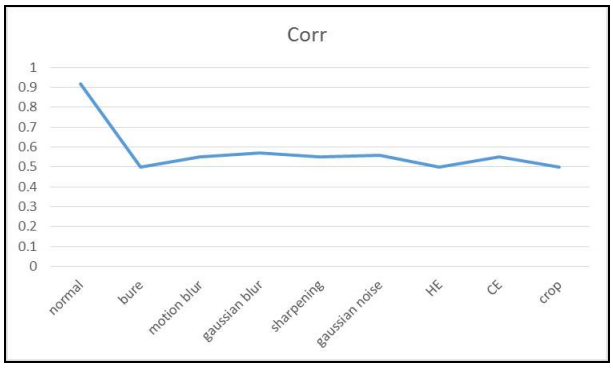


Figure 5: Comparison of different types of modifications done to the watermarked image

Firstly, we add Gaussian noise and try to obtain the watermark image. Our goal is that, inspite of the Gaussian noise we have to get the watermark image. We test the watermarking technique for various levels and types of noise. In the graph given below are the various performance analyses for MSE, PSNR values[12].

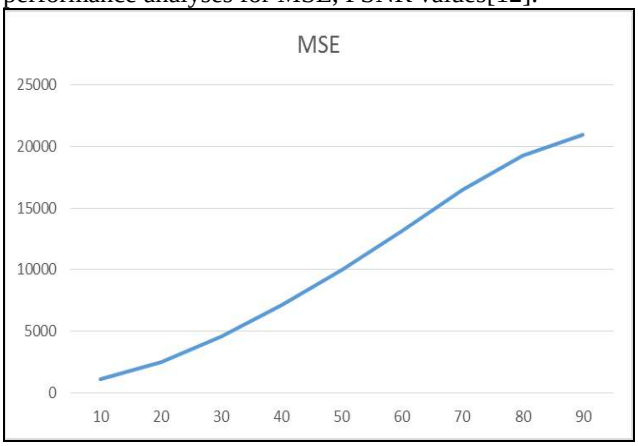


Figure 6: MSE values on adding of Gaussian noise

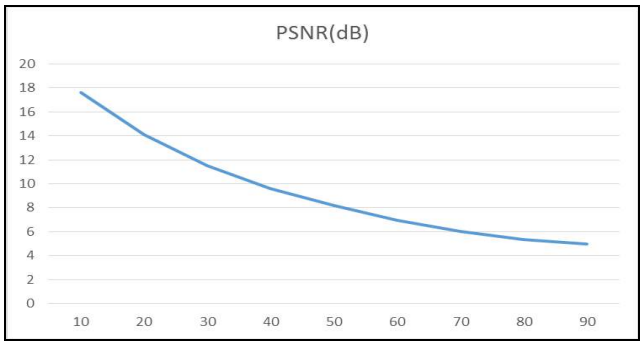


Figure 7: PSNR values on adding Gaussian noise.

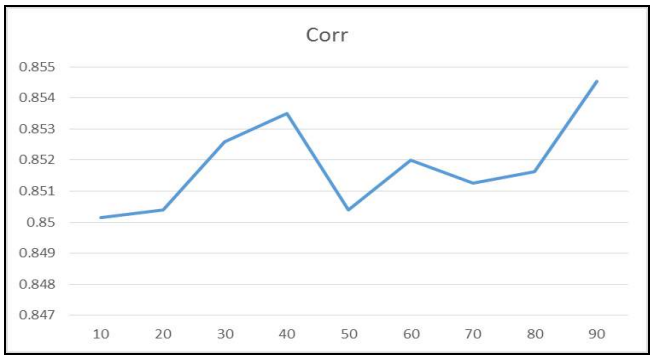


Figure 8: Correlation values for Gaussian noise.

We test the watermarking technique for various levels and types of noise. In the graph given below are the various performance analyses for MSE, PSNR values.

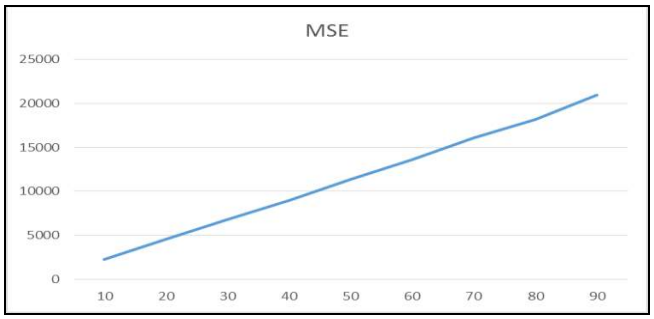


Figure 9: MSE values on adding salt & pepper noise.

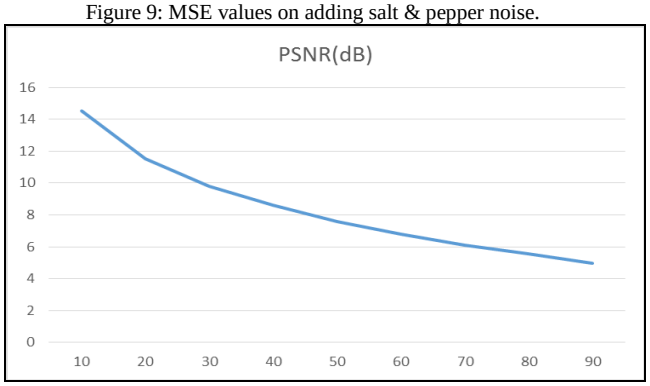


Figure 10: PSNR values on adding salt & pepper noise.

Figure 11: Correlation values for salt & pepper noise.

CONCLUSION

Our actualized DWT-SVD plan has demonstrated a high level of heartiness against lion's share of assaults including solid geometric assaults including editing and different sorts of sign recovering so as to prepare assaults which can be accepted the watermark from any of the sub-band, which plainly shows that change space is more hearty than spatial area. Along these lines, given strategy can be successfully utilized for copyright security of visual data. By and large, LL band is not adjusted as any sort of changes in it can be effectively seen by human eyes. Yet, in DWT-SVD approach, we encountered no such issue.

REFERENCES

[1] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, Importance of Digital Watermarking., in *Digital Watermarking and Steganography*, (USA): Morgan Kaufmann, 2009, 11-12.

[2] S. Dattatherya, V. Chalam, and M. K. Singh, A Generalized Image Authentication based on Statistical Moments of Color Histogram, *International Journal. on Recent Trends in Engineering and Technology*, 8(1), 2013.

[3] M. M. Sathik, S. S. Sujatha, Authentication of Digital Images by using a Semi-Fragile Watermarking Technique, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(11), 2012, 39-44.

[4] M. Ramkumar and N. Akansu, A Robust Protocol for Providing Ownership of Multimedia content, IEEE trans on Multimedia, 6, 2004, 469-478.

[5] E. Ganic, and A. M. Eskicioglu, Robust DWT-SVD Domain Image Watermarking: Embedding date in All Frequencies, *Proc. CiteSeerX, MM&SEC'04*, Magdeburg, Germany, 2004.

[6] P. Havaldar, and G. Medioni, Watermarking Techniques in *Multimedia Systems: Algorithms, Standards, and Industry Practices*, Boston, USA: Course Technology, Cengage Learning, 2010, 414-415.

[7] P. Meerwald and A UbI, A survey of wavelet-domain watermarking algorithms, *Proc. SPIE, Electronic*

Imaging, Security and Watermarking of Multimedia Contents III, San Jose, CA, 2001, SOS-SI6.

[8]C. C. Lai, and C. C. Tsai, Digital Image Watermarking Using Discrete Wavelet Transform and Single Value Decomposition, IEEE Transactions on Instrumentation and Measurement, 59(11), 2010, 3060–3063.

[9]P. Meerwald and A UbI., "A survey of wavelet-domain watermarking algorithms," in *Proc. SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III, vo1.4314*, San Jose, CA, 2001, pp.SOS-SI6

[10] Habibollah Danyali, Morteza Makhloghi, and Fardin Akhlagian Tab, "*Robust Blind DWT based Digital Image Watermarking Using Singular Value Decomposition,"* International Journal of Innovative Computing, Information and Control, Vol. 8, No.7 (A), pp. 4691-4703, July 2012.

[11]http://www.enseignement.polytechnique.fr/profs/informatique/Francois. Sillion/Majeure/Projets/huber/projet.html

[12] http://www.mathworks.in

Enhancing Image Copy-Move Forgery Detection using Particle Swarm Optimization Techniques

MANJESHBN, MADHU BN

Assistant Professor,PG Scholar, Department of Computer Science AIT Tumkur

manjeshbn@gmail.com, madhu.akit@gmail.com

ABSTRACT: Copy-Move Forgery (CMF) is a simple and effective operation to generate forged digital images. Recent, techniques on Scale Invariant Features Transform (SIFT) are commonly used to detect CMF. Various approaches under the SIFT- framework are the most acceptable ways to CMF detection due to the robust performance of SIFT. However, for some CMF images, these approaches cannot produce satisfactory detection results. For instance, the number of the matched keypoints may be too less to prove an image to be a CMF image or to produce an accurate result. These values are only applicable for few images, which limits their application. To solve the problem, a novel approach names as CMF Detection with Particle Swarm Optimizations (PSO) algorithm into the SIFT-based framework. Sometimes these approaches may even produce error results. According to our observations, one of the reasons is that detection results produced by the SIFT-based framework are highly depending on parameters whose values are often determined with experiences. It utilizes the PSO algorithm to generate customized parameter values for images, which are used for CMF detection under the SIFT=based framework. Experimental results show that CMFD-PSO has good performance.

Keywords: CMF: Enhancing copy-move forgery detection: SIFT: region duplication: digital image forensics.

I INTRODUCTION

Copy-Move Forgery (CMF) is a simple and typical operation that tampers with an image by copying at least one part of the image and pasting it to a different location of the same image [6]. The Scale-Invariant Features Transform (SIFT)\ has been proved to have robust performance in detecting this kind of forgery.

Image security is a key issue in any field that makes use of digital images. Images have long been a part of the forensic investigation and law enforcement, an example of which include images of criminals, images of crime scenes, biometric images, etc. However, with the development of the sophisticated techniques for digital image forgery and the low cost to obtain a high-quality digital image, anyone can manipulate a digital image easily without leaving visible clues. Accordingly, digital image forensics has emerged as an important research field.

Unfortunately, all classical CMF detection approaches under the SIFT-based framework have one common drawback, i.e., their detection effects are extremely dependent on the selection of parameter values. In various literature, different parameter values may be seen. Normally, the values are set according to experience or some experiments on a number of forgery images. However, these experience parameter values (EPV) are only applicable to few images. The approaches under the SIFT-based framework, which use the EPV, are names EPV-SIFT in this paper. Sometimes, duplicated regions identified by PEV-SIFT are false while true duplicated regions in a forged image are missed. Sometimes, the number of true matched key points (TMKs) indicated by EPV-SIFT is too less to estimate the duplicated regions accurately. Two examples are shown in Fig.1.

To handle the issue of EPV-SIFT parameters setting, we integrate the Particle Swarm Optimization (PSO) algorithm into the SIFT framework and propose a new approach to detecting CMF. We name our approach as CMF Detection with PSO, or CMFD-PSO, which automatically determines customized parameter values (CPV) for images. With the help of CPV, forged images that cannot be detected by EPV-SIFT will be detected easily. In a word, detection with CPV can obtain much better results than that with EPV.

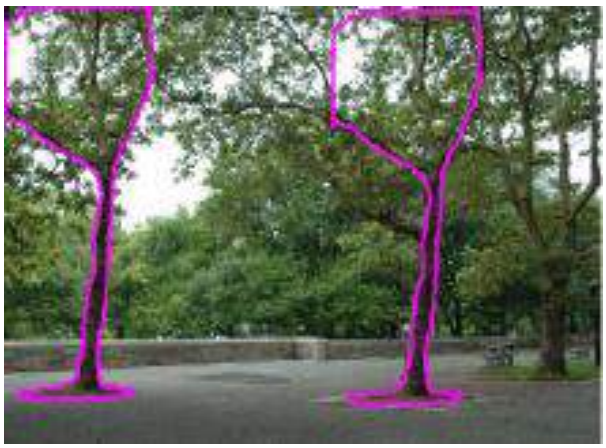
The rest of this paper is organized as follows. Section II introduces related work. Section III analyzes problems

exist in classical SIFT-based CMF detection with EPV. Section IV presents the design of our new approach. Section V devises experiments to test and evaluate our approach. Section VI concludes the paper.

II. RELATED WORK

Several techniques in digital images forensics are employed to detect CMF images. In recent years, approaches under the SIFT-based framework are widely applied to detect based CMF, because they have robust performance in detecting the duplicated regions with geometrical or illumination adjustments. Huang etel. Proposed a preliminary SIFT-based framework in 2008 [3]. They detected SIFT key points and built SIFT descriptors using SIFT algorithm, and then matched these key points to find generally duplicated regions. They noticed that a parameter setting is important for detection results so that they made many experiments find the best parameter value. However, they just found the importance of one parameter but ignored the others.

Amerini et al. [7], Pan and Lyu [6,9], all of them paid high attention to estimating duplicated regions. Although noticing the influence of parameters, they only set specific parameters for their image database. Jing-Ming Guo et al. used DAISY descriptor to detect uniform texture images [10]. There are many efforts similar to Jing-Mingo Guo, which changed some algorithms of theSIFT-based framework to meet some detection purposes.



(a)EPV-SIFT: An error result is shown

(b) Key points to estimate duplicated regions

Fig.1 Images are detected by EPV-SIFT. (a) An error result is shown. (b) The true matched key points are too few to estimate duplicated regions accurately counter-forensics of SIFT-based copy-move detection, the essence of which is to process some key points and make those key points to be ignored by dissatisfying detection condition. However, the detection condition is proposed basing on some parameters. In conclusion, setting parameters is very important for forensics and counter-forensics. With the development of digital images, there are many mature tools that can detect images directly, such an extract image key points, build descriptors, and match key points, etc.

III. FORMULATION OF PROBLEMS

This section analyzes problems in parameter setting after a brief description of the SIFT-based framework.

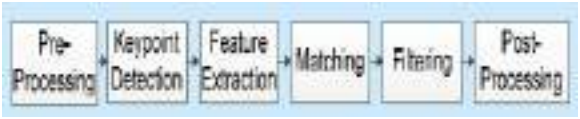


Fig 2: Common workflow of SIFT-based CMF framework

3.1 The SIFT-based framework

Pre-Processing is to prepare an image for detection, such as converting an RGB image into a grayscale image with standard color space conversion.

CMF detection approaches under the SIFT-based framework work in a general waythat may be divided into

Pre-Processing. Keypoint detection.Feature Extraction, Matching, Filtering, and Post-Processing, as shown in Fig 2.

Feature Extraction is to build a descriptor, i.e. a feature vector, for each key point based on its relationship with the surrounding pixels.

Keypoint Detection is to find points that are stable for geometric transformation and illumination transformation as key points.

Filtering is to eliminate mismatch key points, which are identified as matched key points during Matching, bu actually they are not.

Matching is to determine matched key points based on feature vectors. The regions around the matched key points are probably duplicated regions.

Post-processing is to delete duplicated regions, or estimate geometric transformation parameters, and so on, when necessary. It depends on different detection purposes.

The effects of the CMF detection workflow may be shown as Fig. 3.

3.2 Problems in parameter values selection

As detection results depend on the selection of parameter values, an obvious drawback exists in existing CMF detection approaches. Normally, these parameter values are determined by experiences or results of thetest against a number of forgery images. However, different research teams choose different values, which are only applicable to certain images. When they are used to detect a large number of images, the following limitations appear.

1) The number of the matched key points is limited. Using EPVs, there may be very few key points being found in some duplicated regions, or even no key point can be found. In this situation, it is difficult to estimate duplicated regions accurately. To prove that the image is a CMF one with so few matched key points. Some typical examples are shown in Fig 1 (b).
2) Duplicated regions cannot be detected. There are two scenarios. First, the duplicated regions cannot produce key points, or the key points in the duplicated regions are not stable and hence are eliminated in Filtering. Second, no matched key point pairs satisfy the match conditions.
3) Detected regions are not duplicated ones. If there are too many similar objects in an image and parameter values are chosen inappropriately, some similar regions may be mistakenly regarded as duplicated regions, though actually they are native regions in the original image.

IV. DESIGN OF OUR APPROACH

The goal of our approach, CMFD-PSO, is to automatically generate suitable parameter values for each test image. The flow chart of CMFD-PSO is shown in Fig 4. It includes two components, one of which is Elemental Detection.

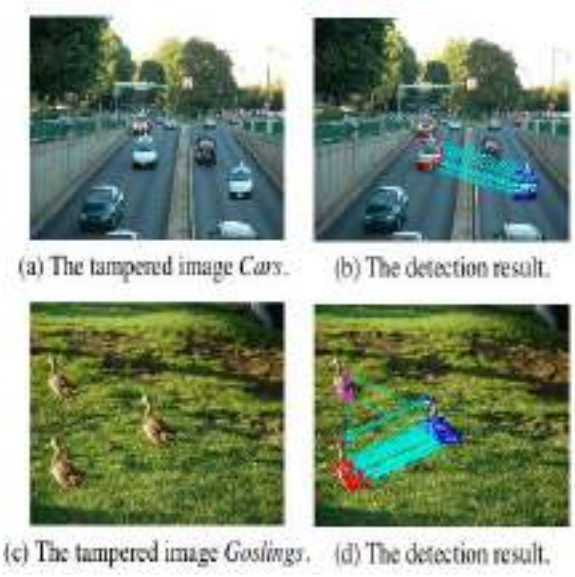


Fig. 3 Detecting CMF under SIFT-based framework

Elemental Detection is derived from the SIFT-based framework. Its task is to detect CMF images. Parameters Estimation is a new component, which can generate suitable parameter values for each image. Using these values to detect the corresponding image may produce a satisfactory result. The PSO algorithm [15,16] is applied to estimate parameter values. To our knowledge, none of the existing CMF detection approaches use the PSO algorithm.

4.1 Overview of CMFD-PSO

CMFD-PSO generates suitable parameter values automatically for each image according to the features of the image. With these parameter values, Elemental Detection can produce better results.

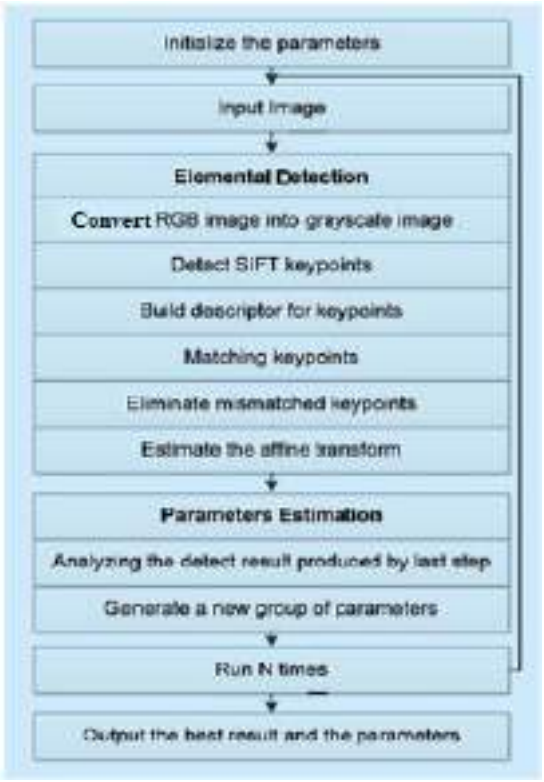


Fig. 4 the flow chart of our approach, CMFD-PSO

The first step is to identify the input and the output of the SIFT-based framework. The input includes an image and a group of parameters. The output is only the number of matched key points, which is used to evaluate whether the results is good. We turn parameter value estimation into an issue of theoptional solution. An evaluation criterion is created to make detection decision. The criterion is formed by the number of matched key points. When the criterion reaches extreme value, theoptional solution will turn out. In PSO, the parameter value estimation issue can be expressed as:

$$D_{Result} = f(X), \quad X = (x_1, x_2, x_3, x_4 \ldots)$$

X is a group of input parameters: f(x) is detection process. Dresult is the detection result. By adjusting the values of X, Dresult can converge to the extreme value.

In this paper, the PSO algorithm is applied to solve the optimal solution problem. The PSO algorithm is proposed by Eberhart and Kennedy in 1996 to model thesocial behavior of bird flocking or fish schooling. The algorithm is suitable for solving minimization or maximization problems.

Usually, there are some solutions for the optimal solution problem. If the method of exhaustion is used, it will cost a lot of time. It may take more than one year to detect one image. It is impractical.

Using CMFD-PSO to detect images, initially, random or manually generated initialization parameter values are used, then, the following two operations are executed N times.

(1) According to the result of theoperation (1), a new group of parameter values is generated by Parameters Estimation. Then deliver this group of parameter values to operation (1) and start the next round.
(2) Elemental Detection detects the input image with the detection parameter values and then delivers the detection result to operation (2).

The best detection result is chosen from the operations of the N rounds. Then this result and relevant parameter values are output. In our experiment, we set the value of N to 100.

4.2 The elemental detection

In Pre-Processing, an RGB image should be converted into a gray-scale image. In Keypoint Detection and Feature Extraction, the key points are detected from the test images and the SIFT descriptor, a 128-dimensional feature vector, is built for the corresponding key points.

This component consists of five steps that are similar to those of the SIFT-based framework, which is shown in Fig 2. The details of each step are shown in the following instructions.

In Filtering, the mismatched key points should be eliminated. If the distance between two matched key points is too small, this pair of matched key points will possibly be a mismatch. The descriptors of such pair of matched key points may be very similar. In this paper, if the s=distance of the paired key points is smaller than a preset value Dts min they will be removed, which can reduce the probability of mismatching key points. The other mismatched key points are eliminated by the Random Sample Consensus (RANSAC) algorithm [8]. Given two matched key points sets from a region and its duplicate as P and P*, respectively. They are related by an affine transform specified by a matrix T and shift P0 vector as P*=TP+P0. The following steps will run M times: Three pairs of non-collinear matched key points are randomly selected to obtain a transform parameter T and shift vector P0. Then, all pairs of matched key points are classified into inliers or outliers. Specifically, a pair of matched key points (P, P*) is an inlier if ||P*-TP-P0|| < R. otherwise, it is an outlier.

In Matching, the best-bin-first algorithm (BBF) [5] is applied to match key points. When looking for matching with feature vector f1, another feature vector f2 should be found according to the smallest Euclidean distance l1 between the vectors. Then, a third feature vector f3 other than f1 and f2 should be found, where the Euclidean distance l2 between f1 and f3 is the second smallest. The match condition is l1 < rl2, where r is a threshold and r E (0, 1).

T and P0 are estimated basing on the number of inliers, choosing the largest number of inliers from M times estimate results. If the matched key points do not meet the conditions that they will be regarded as the mismatch.

4.3 The parameters estimation

The PSO algorithm is used to search the adjustable parameters. The PSO algorithm is suitable for solving minimization or maximization problems. Before using PSO to find customized parameter values, we should endeavor to explicitly answer two questions that inevitably emerge:

(1) How to build the evaluation function to choose the customized parameters?
(2) Which detection parameters of the SIFT-based framework need to be optimized?

4.3.1 Parameters for elemental detection

The parameters of the SIFT-based framework need to be optimized and their boundaries are listed in Table 1. The reason for the choice of these parameters is that these parameters will make anevident effect for final detection results.

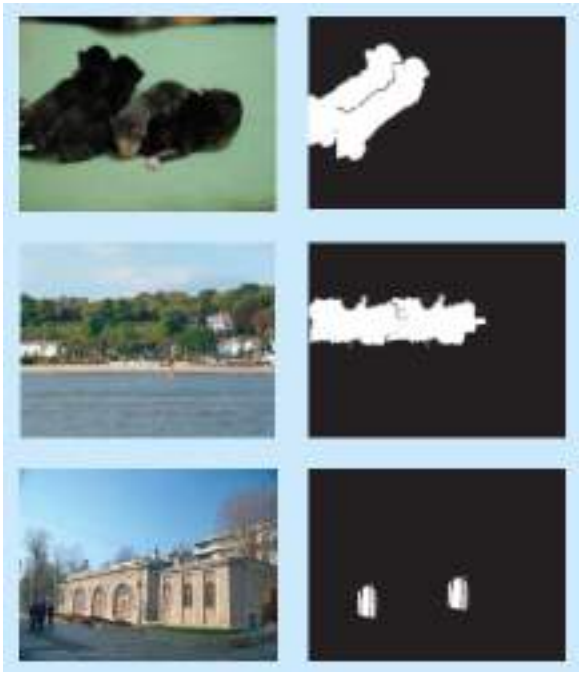


Fig. 5 Examples used in our comparisons

CONCLUSIONS

In this paper, propose a novel approach, CMFD-PSO, to detecting CMF in digital images. Comparing with existing work, thepaper makes three contributions.

(1) It derives rules to automatically determine customized parameter values for given images that are to be detected.
(2) It puts forward the concept of applying the PSO algorithm to CMF detection

(3) It integrates the PSO algorithm into the SIFT-based framework to perform CMF detection.

Fig. 6 Three CMF images that cannot be detected by neither CMFD-PSO nor EPV-SIFT

We proof the concept of CMDF-PSO by experiments. Experimental results show that CMFD-PSO can automatically generate customized parameter values for images, which are independent of neither experiences nor experiments.

Three examples are shown in Fig 6. As a future work, we will figure out new ways to improve the detection performance for such cases. CMFD-PSO can achieve much better results than EPV-SIFT in that it can identifies matched points that its counterparts cannot, and it can dramatically increase the number of true matched key points, which make the detection of region duplication more accurate and more acceptable.

ACKNOWLEDGEMENTS

This work was supported in part by National Natural Science Foundation of China under grant number (61472429, 61070192, and 91018008), Beijing Natural Science Foundation under grant number 4122041. National High-Tech Research Development Program of China under grant number 2007AA01Z414, and National Science and Technology Major Project of China under grant number 2012ZX01039-004.

REFERENCES

- H L Huang, WQ Guo, Y Zhang, "Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm", IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Vol 2, pp 272-276, December 2008.
- O M, Al-Qershi, B E Khoo, "Passive detection of copy-move forgery in digital images: State-of-the-art", Forensic Science International, Vol 231, no 1-3, September 2013.
- D G Lowe, "Distinctive Image Features from Scale-Invariant eypoints", International Journal of Computer Vision, Vol 60, no 2, pp 91-110, November 2004.
- J S Beis, D G Lowe, "Shape indexing using approximate nearest neighbor search in high-dimensional spaces", Computer Vision and Pattern Recognition, San Juan, PR, pp 1000-1006, Jun 1997.
- V Christein, C Riess, J Jordan, C Riess, E Angelopoulou, "An Evaluation of Popular Copy-Move Forgery Detection Approaches", IEEE Trans. Inf. Forensics Security, Vol 7, no 6, pp 1841-1854, December 2012
- X.Y Pan, S.W Lyu, "Region duplication detection using image feature matching",
- I Amerini, L Ballan, R Caldelli, A.D Bimbo, "A SIFT-based forensic approach for copy-move attach detection and transformation recovery".
- X.Y Pan, S.W Lyu, "Detecting image region duplication using SIFT feature"
- R Hess, OpenSIFTrobwhess.github.io/opensift/
- A Vedaldi, B Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms

A Survey on Different Methods for Liver Segmentation

Mallikarjun Kesaratti, Dr. Sunanda Dixit, Prof. B.I. Khodanpur
PGscolar, Associate Professor, Professor,
Dept. of Information Science and Engineering, DSCE, Bangalore
mallikarjun.sk@outlook.com, sunanda.bms@gmail.com, bi.khodanpur@gmail.com

ABSTRACT: Liver Cancer is one in every of the speediest growing cancer in the world. The early detection and diagnosing of liver tumor growth is vital for the hindrance of liver tumor growth. More than 30% of cancer deaths may be prevented by avoiding risk factor, early detection, accurate diagnosis, and effective treatment. Segmentation of liver from medical images from the abdominal space is vital for diagnosing of tumor and for surgical procedures. Accurate detection of the type of the liver abnormality is very essential for treatment designing which may minimize the fatal results. However accurate results can only be obtained by computer aided automation systems. Many different techniques are developed for the detection of liver cancer the abnormal lesion size and form. This paper reviews various liver tumor detection algorithms and methodologies used for liver tumor diagnosis. A comparative analysis is performed. Also explores the applicability of the techniques in liver segmentation of CT images.
Keywords- **Level set Segmentation, K- means clustering, liver cancer, CT image**.

I. INTRODUCTION

Liver tumors or hepatic tumors are tumors or growths on or in the liver. Several distinct types of tumors can develop in the liver because the liver is made up of various cell types. liver cancer tends to occur in livers damaged by birth defects, alcohol, or chronic infection with diseases such as hepatitis B and C, hemochromatosis (a hereditary disease associated with too much iron in the liver) [1].

New liver cancer cases 39,230 adults (28,410 men and 10,820 women) in the United States in 2015. Death due to liver cancer in USA, 2015 is 27,170 deaths (18,280 men and 8,890 women). The survey was done on the basis of survival rate. If the person has survived for 5 years then the survival rate is chosen as 5 years. The first year survival rate for people with liver cancer is 44%. And fifth year survival rate for people with liver cancer is 17%.

Liver cancer is one of the most popular cancer diseases and causes a large amount of death every year. The chances for liver cancer in men and women have increased to 40% and 23% respectively.

Image segmentation is another useful tool used in the field of image processing. the main purpose of the this idea is to capture like characteristics of an image and bring them out to where they are more visible than they were before. Segmentation of images is used to provide information such as structures of organs, identifying the regions of interest i,e locating tumors, abnormalities etc. the liver is the largest gland ant he largest internal organ in the human body, liver is dark red, wedge-shaped gland approximately eight and half inches long. Early detection and accurate analysis of liver cancer is an important issue in practical radiology. Liver lesions are the injury, wound, disease or tumor to the liver tissues. CT scan can identify the liver lesions by difference in pixel intensity from that of the liver. Manual segmentation is very time consuming and tedious process, were as the automatic segmentation is very challenging task due to the factors like indefinite shape of the lesions and low intensity

contrast between lesions and similar to these of nearby tissues. The irregularity in the liver shape,size between the patients. Various automatic/semiautomatic techniques for liver tumor segmentation have been developed.

II. LITERATURE OVERVIEW

R.Rajagopal. et al. [2] proposed a novel system for detecting and segmenting liver lesions. It utilizes a otsu's thresholding method and is employed in median filtering using mathematical morphology. Morphological filtering is applied to extract the regions shap I,e edges. Only erosion is used. And gabor transform filter used for edge detection process. It yields accurate results for different types of liver tumors with ease and without manual interaction. It can also be improved by neural network and fuzzy algorithm.

Gang Chen. et al. [3] multiple initialization, multiple-step LSM are used. The multiple-initialization curves are first evolved separately using the fast marching methods and LSM, which are combined with a convex hall (CH) algorithm to obtain a rough liver contour. Parallel propagation using FMM and LSM based on these initial curves are implemented. Combination of the partial segmentation results using a CH algorithm. Smoothing

the primary liver contour using LSM. Multiple-initialization LSM is much faster, can cover more liver regions. Overcome the leakage and over segmentation problem. An automated perform perfusion analysis method is proposed to automatically conclude liver perfusion curves.the under segmentation problem still exists on lower sharp corner regions due to the low-gradient definition of the lower half of the liver regions in the abdominal MRI's

Chen Zhaoxue. Et al. [4] Simple line search method for plane domain segmentation to extract binary image composed of isolated white pixel clusters mainly from the liver part based on the histogram distribution and spatial characteristics the liver is obtained. Gaussian blurring technique is introduced to connect the isolated pixel clusters. Thresholding The blurred image after the post-processing step of mending holes and size filters. Liver image registration between slices so as to increase accuracy of the perfusion computation and measurement.

S. Luo et al. [5] a three step liver segmentation algorithm. Texture analysis is applied into abdominal CT images to extract pixel level features. Two other main features are wavelet coefficient and haralick texture description are used. here SVM is implemented to classify the data into pixel-wised liver or non liver. Morphological operation is designed as a processor to remove noise and to delineate the liver. It has been proven that wavelets features present better classification the haralick texture descriptors when SVM are used. The combination of morphological operation with a pixel-wised SVM classfier can delineate volumetric liver accurately.

Shraddha Sangewar et al. [6] the segmentation is based on combining a modified K-means segmentation method with a special localized contouring algorithm in the segmentation process in order to divide the image, five separate regions are identified on the T image frames. It provides fast and accurate liver segmentation and 3D rendering as well as delineating tumor regions.

O. Fekry Abd-Elaziz et al. [7] combination of intensity analysis, region growing and pre-processing steps for automatic segmentation of liver and a second region growing process for tumors segmentation. a method for automatic segmentation of liver tumor. Decrease the computation time by removing the regions of other structures. In most techniques liver was segmented using region growing method that started from a seed point automatically selected.

Wenhan Wang et al. [8] a morphological feature of the liver region under various window level setting, applied the region growing algorithm to remove other tissues such as skeleton, kidney & stomach. a discrete points of the liver region can be acquired. The gradient information based edge correction and three dimensional restoration are adopted to optimize the recovered liver image. And has a lower time complexity but there is likely over segmentation.

Ina Singh et al. [9] Discussed he standard k-means clustering algorithm and analyzes the shortcomings of standard k-means algorithm, such as the k-means clustering algorithm has to calculate the distance between each data object and all cluster centers in each iteration, which makes the efficiency of clustering was not high. This paper proposes an improved k-means algorithm in order to solve this question, requiring a simple data structure to store some information in every iteration, which was to be used in the next interation. The improved method avoids computing the distance of each data object to the cluster centers repeatly, saving the running time. Experimental results show that the improved method can effectively improve the speed of clustering and accuracy, reducing the computational complexity of the k-means.

Gambino, O. and et al. [10] proposed an automatic texture based volumetric region growing method for liver segmentation was proposed. 3D seeded region growing was based on texture features with the automatic selection of the seed voxel inside the liver organ and the automatic threshold value computation for the region growing stop condition. Co-occurrence 3D texture features are extracted from CT abdominal volumes and the seeded region growing algorithm was based on statistics in the features space.

III. METHODS FOR TUMOR SEGMENTATION

A.K-Means Clustering

K-means is one of several simplest unsupervised learning algorithms that classify a given knowledge set into bound amount of clusters i,e K-clusters. The idea is to define a K-centroids, one for every single cluster [6]. These centroids must certainly put right into a cunning way, because different location causes different results. So, the best choice is to position them as far as possible from each other. The next issue is usually to require every point belonging to a given knowledge set and associate it to the closest centroid.

When no point is pending, the first step is completed and a early grouping is done. Again re-calculate new K-centroids of the clusters, resulting from the previous step.

After having these k new centroids, a new binding needs to be performed between the same knowledge set points and the nearest new centroid[11]. Repeat the steps until centroids don't move any more. In the successive loops, the k centroids change their location detail by detail.

It is classified as a pixel-based image segmentation technique because it involves the selection of initial seed

points. This technique starts with initial "seed points" and then examines neighbouring pixels (using either 4-connectivity or 8-connectivity) to find out perhaps the pixel neighbours ought to be added with the region. The method is iterated on, within the very same manner as general data clustering algorithms. The region growing algorithm is

Described in as: (i) Select several seed points. Seed point selection is dependent on some user criterion (for example, pixels during a specific gray-level range, pixels equally spaced on a grid, etc The first region begins as the complete precise location of the seeds. (ii) The regions area unit then grownup from these seed points to adjacent points according to a location membership criterion.

The criterion might be pixel intensity, grey level texture or color. Due to the fact the regions are grown on the building blocks of the criterion, the image data itself is very important. For instance, if the criterion were pixel intensity, examine all the adjacent pixels of seed points. If they are of the same intensity value with the seed points, classify them to the seed points. it's for sure an iterated process until there's no modification in two successive iterative stages. The appropriate alternative of seed points is simply a major issue.

B. Level set method

This methodology will handle topological changes and outline the problem is in higher dimension, however this technique is time overwhelming and leads to over segmentation [3]. The Segmentation using level set method this evolves according to a speeding image that is the results of a scanning technique based dynamic programming. The main limitations of this level set method adjust this first segmentation using a speed function which is obtained from the pixel classification algorithm.

Table 1: Survey of different Segmentation

Ref No.	Authors	Year	Technique used	Advantages	Limitations
2	R.Rajagopal	2014	Otsu's thresholding , morphological filtering and gabar transform	Accurate results are yields for different types of liver tumors	Can be improved by using neural network and fuzzy algorithm.
7	El-Masry W.H	2014	Invasive Weed Optimization	multi-objective optimization	Computational time complexity is high
8	O.Fekry Abd-Elaziz	2014	Intensity analysis, Region Growing and morphological processing	Decreases the computation time by removing the regions of other structures.	
6	Shraddha Sangewar	2013	Modified K-means with special contouring algorithm	Provides fast and accurate liver segmentation	
11	Amir H. Forouzan	2013	K-Means Clustering and Geodesic Active Contour	Effective Segmentation in Low contrast Images	Used only for low contrast images
10	Wenhan Wang	2012	Region growing algorithm	Lower time complexity	Over segmentation
5	Zhaoxiao Yuan	2010	Fast Marching and Improved Fuzzy cluster Method	Accurate Segmentation of abdominal Images	Over segmentation
9	O. Gambino	2010	Texture Based region growing	Accurate Recognition and segmentation	Processing time is high
4	Suhuai Luo	2009	Texture analysis and combined morphological operations	the combination of morphological operations with pixel-wised SVM classifier can delineate volumetric liver accurately	
3	Gang Chen	2009	Level set method	Multiple-initialization LSM is much faster, covers more liver regions, overcome the leakage and over segmentation.	LSM is the difficulty in automatically determining the number of initialization, under segmentation still exists on lower sharp corner regions.

The accuracy here is only sufficient in a small number of cases. The level set method is initially proposed for tracking moving interfaces and has spread across the various imaging domains. It can be used to efficiently address some problems such as curve/surface/etc. propagation in an implicit manner. The central idea is to represent the evolving contour by using a signed function whose zero corresponds to the actual contour. Then, the motion equation of the contour, one can derive a similar flow for the implicit surface that when applied to the zero level that will reflect the propagation of contour. The level set method affords numerous advantages: it is implicit, is parameter free, provides a way to estimate the geometric properties of the evolving structure, allows for a change of topology, and is intrinsic. It is used to define

an optimization framework. One can conclude that it is a very convenient framework for addressing various applications of computer vision and medical image analysis. Research in various level set data structures has led to very efficient implementations of this method.

IV. CONCLUSION

In this paper, a survey on various image segmentation techniques has been done. the K-Means algorithm is limited to only initial number of k-clusters. Wrongly selected clusters lead to erroneous results Therefore in future, level set method for segmentation is suggested which will overcome the problem of K-Means clustering.

REFERENCES

[1]Cancer research UK," http://www.cancerresearchuk.org/".

[2] R. Rajagopal , P. Subbaiah "A Survey On Liver Tumor Detection And Segmentation Methods", *ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 6, APRIL 2015*.

[3] G. Chen, L. Gu , L. Qian , J. Xu "An Improved Level Set for Liver Segmentation and Perfusion Analysis in MRIs",*IEEE Transactions on Information Technology in Biomedicine, Volume:13 , Issue: 1, 31 October 2008.*

[4] Chen Zhaoxue,Nie Shengdong, Qian Lijun,Chen Zeng'ai, Xu Jianrong," Automatic Liver Segmentation Method Based on a Gaussian Blurring Technique For CT Images" *IEEE Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on 18 May 2008.*

[5] Suhuai Luo, Jesse S. Jin, Stephan K. Chalup, Guoyu Qian,"A Liver Segmentation Algorithm Based on Wavelets and Machine Learning",*IEEE 2009*

International Conference on Computational Intelligence and Natural Computing 7 June 2009.

[6] Shraddha Sangewa, Atish A.peshattiwar, Vilas Alagdeve,Rupali Balpande,"Liver segmentation of CT scan images using k-means Algorithms", *2013 Internation Conference on Advanced Electronic System(ICAES).*

[7] O. Fekry Abd-Elaziz, M. Sharaf Sayed, M.Ibrahim Abdullah," Liver Tumors Segmentation from Abdominal CT Images using Region Growing and Morphological Processing", *IEEE Engineering and Technology (ICET), 2014 International Conference on 19-20 April 2014.*

[8] Wenhan Wang, Xihe Gao," A High-speed Method for Liver Segmentation on Abdominal CT Image", *IEEE Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on 23-25 Oct. 2012*

[9] Ina Singh, Asst. Prof. Neelakshi Gupta,"Segmentation of Liver using Hybrid K-means Clustering and Level Set", *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 8, August 2015.*

[10] O. Gambino and et al., " Automatic Volumetric Liver segmentation using Texture based region growing", *Complex ,Intelligent and Intensive Systems, 146-152, IEEE 2011.*

[11] H.P. Ng and et al., " Medical image Segmentation using K-Means clustering and improved watershed algorithm" *Image analysis and interpretation, 61-65, IEEE 2006.*

[12] Zhaoxiao Yuan and et al., " A novel automatic liver segmentation technique for MR images", *Image and Signal Processing, Vol. 3, 1281-1286, IEEE 2010.*

[13] El Masry and et al., " Automatic liver CT image clustering based on invasive weed optimization algorithm", *Engineering and Technology, 1-5, IEEE 2014.*

Certain investigations on filter performance for skin texture analysis

Dr.Punal M Arabi, Gayatri Joshi, Kavya P Parameswaran, Abhik Raj Subedi
Department of Biomedical Engineering, ACSCE, Bangalore-74
arabi.punal@gmail.com, gayitrijoshi@gmail.com , meet.kavya007@gmail.com, abhikbook@gmail.com

ABSTRACT: Noises are always present in an image posing higher or lower complexity in removal and however it is necessary to remove those noises to obtain a better image. There are various types of filters available that help us to remove those noises; only when a right type of filter is used the best result could be obtained. This paper compares the performance of filters namely median, Weiner, average and sharpened filter applied for skin texture analysis by using GLCM(Gray level co-occurrence matrix) parameters. Cheek and dorsal skin images at two different angles are taken by a 8MegaPixel camera, with a resolution of 1920x2560, exposure time $1/10^{th}$ of a second as samples for analysis.

Keywords - Skin texture analysis, GLCM,Filters

I. INTRODUCTION

Various skin texture patterns appear on the surface of human skin. For instance, a person's palm shows different textures of blood spots, prints and wrinkles. The analysis and synthesis of these texture patterns is important for several research areas, such as computer graphics, medical imaging, and cosmetics development. In signal processing, a filter is a device or process which removes the unwanted component or feature from the signal. Filtering is a class of signal processing. Most often, this means removing some frequencies and not others in order to suppress interfering signals and reduce background noise.

II. LITERATURE SURVEY

ShijieHao.etal[1] constructed a spatially guided map which exploits the spatial influence of the image details based on edge response of an image. Further, this map was integrated into two state of the art image filters for image detail enhancement. We know noise removal is important task in different applications such as medical which the noise free images could lead to less error detection. Azadeh Noori Hoshyar.etal [2]compared the performance of five filters - Median Filter, Adaptive Median Filter, Mean Filter, Gaussian Filter and Adaptive Wiener filter- for de-noising from Gaussian noise, Salt & Pepper noise, Poisson noise and Speckle noise.MotonoriDoi etal[3] Proposed image analysis and synthesis of skin color texture by wavelet transform. The skin color texture is modeled as four texture components of base color, internal skin texture, regular surface texture and local texture. The skin color image is decomposed to the four texture components by multi-resolution analysis using wavelet transform. Sergey Abramov etal[4]. proposed Prediction of Filtering Efficiency for DCT-based Image Denoising.It is possible to estimate the MSE values of images to be processed by means of calculation rather simple statistics of DCT coefficients. Moreover, the quasi-optimal value of threshold parameter for DCT filtering methods can be easy evaluated as well. The results are presented for different additive Gaussian noise levels and a set of gray-scale test images. Ji-Hong Liu etal[5] proposed Research and Implementation for Texture of Handback Skin Quantitative Analysis based on Co-occurrence Matrix. First of all, the non-invasive method of digital imaging Technology is adopted to get the handback skin morphological data. Then, based on the algorithms of gray level co-occurrence matrix and displacement co-occurrence Matrix, morphological features of the individual hand back skins are measured. Finally, we analyses the relationship between the experimental results and skin aging for both methods.Kouhei Shimizu.etal[6] proposed a new computer-aided method for the skin lesion classification applicable to both melanocytic skin lesions (MSLs) and nonmelanocytic skin lesions (NoMSLs).They developed a new method to distinguish among melanomas, nevi, BCCs, and SKs. They calculated 828 candidate features grouped into three categories: color, sub region, and texture. They introduced two types of classification models:a layered EmreCelebi.etal[7],proposed the automatdquantification o f clinically significant colors in dermoscopy images.
Given a true-color dermoscopyimage with N colors, we first reduce the number of colors in this image to a small number K, i.e., K <; N, using the K-means clustering algorithm incorporating a spatial term. The optimal K value for theimage is estimated separately using five commonly used cluster validity criteria.Mariam A.Sheha.etal[8], presented an
automated method for melanoma diagnosis applied on a set of dermoscopy images. Features extracted are based on gray level Co-occurrence matrix (GLCM) and Using

Multilayer perceptron classifier (MLP) to classify between Melanocytic Nevi and Malignant melanoma. MLP classifier was proposed with two different techniques in training and testing process.

In this work tells about the novel method of identifying the filter performance of different filters on skin texture analysis.

III. METHODOLOGY

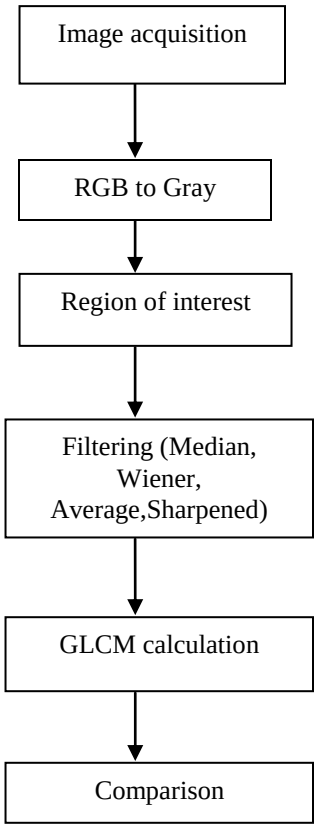


Fig1:Flow diagram

Cheek and dorsal skin images at two different angles are taken by a 8MegaPixel camera, with a resolution of 1920x2560, exposure time 1/10th of a second as samples for analysis. The images are converted to gray.The region of interest (ROI)is then selected. The ROI must have at least 600 to 800 pixels to get reliable results. Different filters like Median filter, Wiener filter, Average filter, Sharpened filter are applied to ROI of these images and the GLCM values are then calculated. The calculated GLCM values of the filtered images by different filters have been compared to analyze the performance of them.

GLCM matrix has been found using the formulae using below:

Contrast $= \sum_{(i,j)} |i-j|2\ p(i,j)$ ..... (1)

Where p=image, i,j=coordinates , p(i.j)=Intensity value at i,j ,

Correlation = $\sum_{i,j} \frac{(i-\mu i)(j-\mu j)\,p(i,j)}{\sigma i - \sigma j}$ ..... (2)

Energy = $\sum_{i,j} p(i,j)2$ .....(3)

Homogeneity = $\sum_{i,j} \frac{p(i,j)}{1+(i-j)}$ .....(4)

Medianfilter:

$y(m,n)=\text{median}\{x[i,j],(i,j)\ \varepsilon\ \}$ .....(5)

where w represents a neighborhood defined by the user, centered around location [m,n] in the image

Wiener filter:

$$f_{ij}=\sum_{K=-n/2}^{(n/2)-1} \sum_{l=-n/2}^{(n/2)-1} (v_{kl} g_{i+k,j+l}) + e_{ij} \quad \text{.....(6)}$$

for I,j=1,2,3,4….n.

where v denotes the weights by which g is blurred and e denotes the noise

Average filter:

$y[i]=\frac{1}{M}\sum_{j=0}^{M-1} x[i+j]$ .....(7)

where x[] is the input signal, y[] is the output signal and M is the number of points in the average

Sharpened filter:

$S_{ij}=x_{i,j}+\lambda F(x_{i,j})$ .....(8)

Where $x_{i,j}$ is the original pixel value at the co-ordinate I,j , F is the high pass filter, λ is the tuning parameter

IV. RESULTS

Fig2:Cheek and dorsal skin images in two different angles

Table1:The GLCM values of Different filters

S. No	Position	Parameters	Filters			
			Median filter	Wiener filter	Average filter	Sharpened filter
1	Hand (angle1)	Contrast	0.1377	0.1846	0.1687	5.6582
		Correlation	0.9445	0.7743	0.7827	0.0955
		Energy	0.1885	0.3405	0.3616	0.0260
		Homogeneity	0.9315	0.9110	0.9189	0.4755
2	Cheek (angle1)	Contrast	0.1176	0.1931	0.1562	6.2551
		Correlation	0.7803	0.7743	0.8230	0.0860
		Energy	0.3095	0.3119	0.3206	0.0231
		Homogeneity	0.9121	0.9074	0.9244	0.4532
3	Hand (angle2)	Contrast	0.1193	0.1488	0.1280	3.5709
		Correlation	0.9478	0.9432	0.9435	0.3979
		Energy	0.2016	0.1852	0.2015	0.0282
		Homogeneity	0.9407	0.9262	0.9380	0.5248
4	Cheek (angle2)	Contrast	0.1302	0.2143	0.1476	8.3461
		Correlation	0.7625	0.6488	0.7666	0.0452
		Energy	0.4564	0.3500	0,4169	0.0192
		Homogeneity	0.9532	0.9025	0.9296	0.4238

Fig3: Cheek skin images by different filters

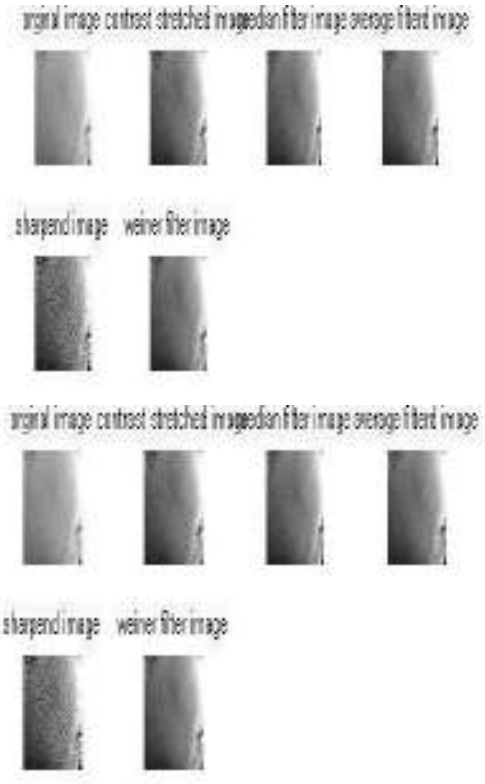


Fig4:dorsal skin images using different filters

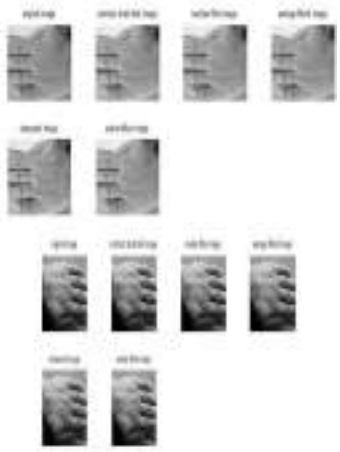


Figure 2 shows the cheek and dorsal hand skin images and figure 3 shows the filtered cheek images taken two different angles. Figure 4 shows the filtered hand images taken at two different angles. Table1 shows the GLCM(gray level Co-occurrence matrix) values of the cheek and hand skin images using different filters.

V. DISCUSSION

The cheek and hand samples are taken for experimentation and imaged at two different angles. A region of interest is identified in each sample and various filters are used to filter this region of interest. GLCM parameters are found for these filtered images.

The results which has been tabulated in Table-1 shows the GLCM(gray level Co-occurrence matrix)values for Median, wiener, Average and sharpened filters. The GLCM values of median, wiener and average filters are not of any particular pattern but only the GLCM values of sharpened filter is of particular pattern for the skin samples taken ; i.e., for median filter the contrast value for hand and cheek at angle1 is decreasing and for angle2 it is increasing. Whereas the correlation for hand and cheek at angle1and angle2 is been decreasing. The energy for hand and cheek at angle1 is and angle2 is increasing. The homogeneity for hand and cheek at angle1 is decreasing, whereas at angle2 it is increasing. It is found that the results of median filter do not yield any particular pattern by comparing all the GLCM values at different angles. For wiener filter, the contrast for hand and cheek at both angle1 and at angle2 is increasing. The correlation for hand and cheek at anlgle1 is found to be the same where as at angle2 it is decreasing; the energy for hand and cheek at angle1 is decreasing and at angle2 is increasing. The homogeneity for hand and cheek at angle1and angle2 is found to be decreasing. Also wiener filter, average filter do not give any particular pattern but only sharpened filter shows a definite pattern; i.e., the contrast for hand and cheek at angle1 and angle2 is increasing. The correlation

for hand and cheek at angle1 and angle2 is decreasing. The energy for hand and cheek at angle1 and angle 2 is found decreasing. The homogeneity for hand and cheek at angle1 and at angle 2 is increasing. From the tabulated GLCM results it is found that, of all the filters considered here only sharpened filters results fall in a pattern for analyzing skin texture.

VI. CONCLUSION

The proposed method is tested on the cheek and dorsal skin images taken at two different angles. The results obtained show that the sharpened filter performs better compared to the other filters namely Median, wiener and average filters to classify the skin textures based on GLCM values. It is found that the contrast of Hand skin is lesser than that of cheek skin where as correlation, Energy and homogeneity of Hand skin are more than that of cheek skin.

However more samples of skin from different locations are to be tested to know about the accuracy of the proposed method.

ACKNOWLEDGMENT

The authors thank the Management and Principal of ACS College of engineering, Mysore road, Bangalore for permitting and supporting us to carrying out this research work.

REFERENCES

[1].ShijieHao,DaruPan,YanrongGuo,RichagHong, MengWang ," Image detail enhancement with spatially guided filters", Signal Processing120(2016)789–796, Elsevier journal.

[2]. Azadeh Noori Hoshyar , Adel Al-Jumailya, Afsaneh Noori Hoshyarb," Comparing the Performance of Various Filters on Skin Cancer Images", International Conference on Robot PRIDE 2013-2014 - Medical and Rehabilitation Robotics and Instrumentation, ConfPRIDE 2013-2014, Procedia Computer Science 42 (2014) 32 – 37

[3]. Damanpreet Kaur, Prabhneet Sandhu," Human Skin Texture Analysis using Image Processing Techniques", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064,

[4]. Sergey Abramov, Sergey Krivenko, Alexey Roenko, Vladimir Lukin," Prediction of Filtering Efficiency for DCT-based Image Denoising", 2nd Mediterranean Conference on Embedded Computing , MECD - 2013

[5]. Jaruwan Toontham, Wirat Rattanapitak and Somkait Udomhunsakul,"Comparative Efficiency of Wavelet Filters for Multi-focus Color Image Fusion", 2nd International Conference on Education Technology and Computer (ICETC), 2010.

[6]. Ji-Hong Liu, Qian Gao and Yang Liu," Research and Implementation for Texture of Handback Skin Quantitative Analysis based on Co-occurrence Matrix", International Conference on Industrial Mechatronics and Automation, 2009.

[7].Rashi goel and saranjeet singh,"Skin cancer detection using glcm matrix analysis and svm classifier", international journal of applied engineering and technology issn: 2277-212x (online) an open access, online international journal available at http://www.cibtech.org/jet.htm vol. 5 (1) january-march 2015, pp.6-11

[8].Mariam A.Sheha,Mai S.Mabrouk,Amr Sharawy,"Automatic Detection of Melanoma Skin Cancer using Texture Analysis", International Journal of Computer Applications (0975 – 8887), Volume 42– No.20, March 2011

APPROACHES TO CONTENT-BASED IMAGE RETRIEVAL

Lavanya B, Meghashree M D, Sanjana P,Usha M, Poonam kumari
Bachelors of Engineering Students, Assistant Professor, Computer Science and Engineering Department, VTU
RAJARAJESWARI COLLEGE OF ENGINEERING, BENGALURU-560074
lavanya.lavu2395@gmail.com , Meghashree826@gmail.com, Sanju.kiran27@gmail.com, usham4884@gmail.ccom, Poonam.rrce@gmail.com,

ABSTRACT- The great interest in research on content – based image retrieval has covered the way for a large number of new techniques and systems. The current decade related to image retrieval and automated image annotation, spanning 120 references key contributions are discussed in this paper. We terminate with a study on the trends in volume and impact of publications in field with respect to venues/ journals and subtopics.

Keywords: Annotation, concept, content –based image retrieval.

I. INTRODUCTION

We learned that inherent is a key to progress without loss of what we already possess. Man has traditionally outperformed machines for most tasks while organizing image. Ambitious attempts have been made to make machines learn to understand index and annotate images representing a wide range of concepts with much progress. The requirement is to establish how image retrieval technology can reach out to common man in the same way text retrieval technical do. For eg: google ™ and yahoo!® are household names today, primarily due to benefits raped through their use. We visualize that image retrieval will enjoy a similar success story if concerted effort is made by the research and user communities in that directions without understanding the context .a text based search engine can successfully retrieve documents, there is no easy way for a user to give a lower level description of what image she is looking for.

II. NEW IDEAS AND APPROACHES

There is a continuous effort towards content of image understanding as there is no universally acceptable algorithm for characterizing human vision.

2.1 FEATURE EXTRACTION

To obtain global image features like color histogram or local descriptors like shape and texture features extraction in the preprocessing step in most systems. To show better yield retrieved than uniform quantization and vector quantization with squared error GMVQ is used and also extracts color histograms. Browsing and retrieval suited histogram based descriptors, dominant color descriptors, spatial color descriptors and texture descriptors rigorously tested for inclusion in the MPEG-7 std. Irrelevant shape features and noisy can be removed using contour simplification as the shape is a key attribute of segmented image regions. Shape content which is a new shape descriptor for shape matching proposed with dynamic programming approach. Problem with this approach is that computation of Fourier descriptors and moments is slow, although pre-computation may help produce real-time results. Shape estimates are meaningless without reliable segmentation. Normalized cuts criteria is one approach, even though problem of segmentation in the context of human perception is far from being solved and it is based on spectral clustering and it is extended to textured image segmentation.

2.2 APPROACHES TO RETRIEVAL

Here we concentrate on some of the more recent approaches semantic categorization allows image matching and import aspect of this is its retrieval speed. Hence A semantics- sensitive approach to content based image retrieval has been proposed. Anchoring based image retrieval has been proposed among other new approaches, which is based on fairly intuitive idea of finding a system of representative "anchor" images and deciding semantic proximity between arbitrary image pair in terms of their similarity to these anchors. To improve interface design, visualization and result preprocessing clustering has been applied to image retrieval.

2.3 ANNOTATION AND CONCEPT DETECTION

Automatic concept recognition from visual features of images is more challenging field. This challenge is due to the semantic gap that exists between low level visual features and high level concepts.

About concepts and annotation: To discovery images pertaining to a given concept in the absence of reliable metadata is the primary purpose of content based image retrieval. Automated annotation tends to be more practical for large data sets than a manual process, because it allows for image search through the use of text. Research in the text

domain is the inspiration for many approaches to image annotation.

2.4 RELEVANCE FEEDBACK AND LEARNING

Relevance feedback is a query modification technique, originating in information retrieval, which attempts to capture the user's precise needs through iterative feedback and query refinement. The user's feedback provides a way to learn case-specific query semantics. Usually user's relevance feedback results in only a small number of labeled images pertaining to each high level concept. To overcome the problem of learning from small training sets, a discriminate-EM algorithm is proposed to make use of unlabeled images in database for selecting more discriminating features. The problem with relevance feedback is that after every round of query should be recomputed using a modified similarity measure. The second problem is user's patience in supporting multi-round feedback. One way to reduce the user's interaction is to incorporate logged feedback history into current query.

2.5 HARDWARE AND INTERFACE SUPPORT

Real-world application demand real-time response. The focus is generally on retrieval and annotation performance, presentation of result often takes a back-seat. Subjectivity in the needs as well as interpretation of results is an issue. One way is to allow for greater flexibility in querying/visualization. In image annotation systems, a way to conveniently create sufficiently representative manually annotated training databases is by building interactive, domain games. Designing querying/visualization for image retrieval system, it helps to understand factors like how people manage their digital photographs or frame their queries for visual art images. User studies on various ways of arranging images for browsing purposes is conducted and observation is both visual feature based arrangement and concept-based arrangement have their own merits and demerits.

III. SHAPE FEATURE

Shape is an important visual feature and it is one of the basic features used to describe image content. The shape extracted from the image partially represents projected object.

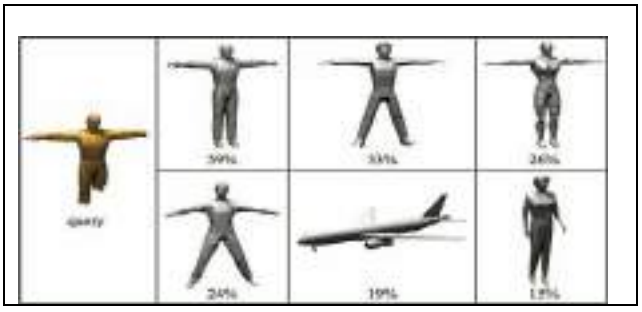


Figure 2: Image search based on body shape

Current approaches have both positive as well as negative attributes. Shape-based image retrieval consists of measuring similarity between shapes represented by their features. Simple geometric features are used to describe shapes. Simple geometric features can discriminate shapes with large differences; thus they are used as filters for eliminating false hits or combining with other shape descriptors to discriminate shape. Shape can be described by different aspects. The shape parameters are mass, centroid, mean, variance, dispersion, eccentricity, circularity ratio, rectangularity, convexity, solidity, Euler number, profiles etc.

IV. REAL WORLD REQUIREMENT

User feedback plays important role I building real-world system during development process. Since there is a so much interest in the field of image retrieval, there is a chance that CBIR based real-world systems will diversify and can even expand feature. Screen shots can be seen in fig-2.there are some issues with implementing CBIR systems on real-world data for public usage. Issues are listed below.

Performance: the current effort is concentrate on improving performance in terms of their precision and recall.

Semantic learning: learning image semantics from training data and developing retrieval mechanism to efficiently leverage semantic estimation are important direction to tackle the problem of semantic gap faced by CBIR.

Figure 1:Image search on nature.net

Volume of data: the software system must be able to efficiently handle indexing and retrieval when the public image databases tend to grow into unwieldy proportions.

Heterogeneity: if the originates from diverse sources causes variation in color and texture features extracted. By

tackling these variations the system can be made more robust.

Concurrent usage: In on-line image retrieval system there will be multiple concurrent users. While most systems have high resource requirement for feature extraction and indexing. They must be efficiently designed so as not to exhaust the host server resources.

Multi-model: The image content can be understood by the presence of reliable meta data such as audio or text caption associated with the images.

User-interface: Effort is needed to the ability to understand the interface for image retrieved so that the people are able use the tool for their profit.

V. CONCLUSIONS

We have compiled research trends in Content Based Image Retrieval and automated annotation using Google scholar's searched tool citation found and exciting fields of CBIR lead us to present a work on it. Here feature extraction and relevance feedback have received a lot of attention, application-oriented aspects such as interface, visualization, scalability and evaluation have traditionally received lesser consideration.

VI. ACKNOWLEDGEMENT

The documentation and implementation of this paper would not be succeeded without the kind support from Dr.R Balakrishna, principal, RajaRajeswari College Of Engineering, Bengaluru, who always gives us valuable advice and kind assistance to complete this paper. We are grateful to Dr. Usha Sakthivel, Professor and HOD, Department Of Computer Science and Engineering, RajaRajeswari College Of Engineering, Bengaluru,for giving us the great knowledge.

Finally, we would like to thank our parents who constantly supported us since the beginning till the end of the paper.

VII . REFERENCES

[1] L. von Ahn and L. Dabbish, "Labeling Images with a Computer Game," *Proc. ACM CHI*, 2004.

[2] Airliners.net, http://www.airliners.net.

[3] J. Amores, N. Sebe, P. Radeva, T. Gevers, and A. Smeulders, "Boosting Contextual Information in Content-Based Image Retrieval," *Proc. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*, 2004.

[4] J. Assfalg, A. Del Bimbo, and P. Pala, "Three-Dimensional
Interfaces for Querying by Example in Content-Based Image
Retrieval," *IEEE Trans. Visualization and Computer Graphics*, 8(4):305–318, 2002.

[5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M.
Blei, and M. I. Jordan, "Matching Words and Pictures," *Journal of Machine Learning Research*, 3:1107-1135, 2003.

[6] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate
Retrieval of Shapes Using Phase of Fourier Descriptors and
Time Warping Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.

[7] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and
Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522,
2002.

[8] D. M. Blei and M. I. Jordan, "Modeling Annotated Data,"
Proc. ACM Conference on Research and Development in Information Retrieval, 2003

[9] C. Bohm, S. Berchtold, and D. A. Keim, "Searching in High-Dimensional SpacesIndex Structures for Improving the
Performance of Multimedia Databases," *ACM Computing Surveys*, 33(3):322-373, 2001.

[10] J. Carballido-Gamio, S. Belongie, and S. Majumdar "Normalized Cuts in 3-D for Spinal MRI Segmentation," *IEEE Trans. Medical Imaging*, 23(1):36–44, 2004.

[11] G. Carneiro and N. Vasconcelos, "Minimum Bayes Error
Features for Visual Recognition by Sequential Feature Selection and Extraction," *Proc. Canadian Conference on Computer and Robot Vision*, 2005.

[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using
Expectation-maximization and Its Application to Image Querying," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(8):1026-1038, 2002.

[13] A. Chalechale, G. Naghdy, and A. Mertins, "Sketch-Based
Image Matching Using Angular Partitioning," *IEEE Trans.
Systems, Man, and Cybernetics*, 35(1):28–41, 2005.

[14] E. Y. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA:
Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines," *IEEE Trans. Circuits and Systems for Video Technology*, 13(1):26–38, 2003.

[15] C.-c. Chen, H. Wactlar, J. Z. Wang, and K. Kiernan, "Digital Imagery for Significant Cultural and Historical Materials - An Emerging Research Field Bridging People,

Culture, and Technologies," *International Journal on Digital Libraries*, 5(4):275–286, 2005.
[16] J. Chen, T.N. Pappas, A. Mojsilovic, and B. Rogowitz,
"Adaptive image segmentation based on color and texture,"
Proc. IEEE International Conference on Image Processing,
2002.
[17] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature
Matching Approach to Content-Based Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):252–1267, 2002.
[18] Y. Chen and J. Z. Wang, "Image Categorization by Learning and Reasoning with Regions," *Journal of Machine
Learning Research*, 5:913-939, 2004.
[19] Y. Chen, J. Z. Wang, and R. Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," *IEEE Trans. Image Processing*, 14(8):1187–1201,
2005.
[20] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5), 603–619,
2002.
[21] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas,
and P. N. Yianilos, "The Bayesian Image Retrieval System,
PicHunter: Theory, Implementation, and Psychophysical Experiments," *IEEE Trans. Image Processing*, 9(1):20–37,
2000.
[22] A. Csillaghy, H. Hinterberger, and A.O. Benz, "Content
Based Image Retrieval in Astronomy," *Information Retrieval*, 3(3):229–241, 2000.
[23] S. J. Cunningham, D. Bainbridge, and M. Masoodian,
"How People Describe Their Image Information Needs: A Grounded Theory Analysis of Visual Arts Queries," *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2004.
[24] C. Dagli and T. S. Huang, "A Framework for Grid-Based
Image Retrieval," *Proc. IEEE International Conference on
Pattern recognition*, 2004.
[25] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and
H. Shin, "An Efficient Color Representation for Image Retrieval," *IEEE Trans. Image Processing*, 10(1):140–147,
2001.
[26] M. N. Do and M. Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," *IEEE Trans. Image Processing*,
11(2):146–158, 2002.
[27] A. Dong and B. Bhanu, "Active Concept Learning for
Image Retrieval in Dynamic Databases," *Proc. IEEE International Conference on Computer Vision*, 2003.
[28] Y. Du and J. Z. Wang "A Scalable Integrated Region-Based Image Retrieval System," *Proc. IEEE International Conference on Image Processing*, 2001.
[29] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth.
Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Seventh European
Conference on Computer Vision, pages 97–112, 2002.
[31] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, andR. Jain, "Content-Based Image Retrieval at the End of theEarly Years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
[30] Global Memory Net, http://www.memorynet.org.

A Survey on usage of FMRI technique in BCI

Gireesh H R[1], Dr.B.S.Pradeep2
[1] VTU Research Scholar, Bengaluru,[2] Director, International R & D dept.,Linyi Top N/w Co. ltd., Shandong, China
gireeshsit@gmail.com, pradeepbs78@yahoo.com,

ABSTRACT- This paper mainly focuses on how Functional Magnetic Resonance Imaging (FMRI) technique will helps in the brain computer interface (BCI). Since BCI uses the neurophysiologic signals originating in the brain to communicate with the devices or computer, FMRI records noninvasively activity of entire brain with a high spatial resolution. Finally, FMRI-BCI provides a novel approach for studying the brain regions, which in turn helps the communication. Technological advancement in higher field Magnetic Resonance Imaging (MRI) scanners, fast data acquisition sequences, preprocessing algorithms, and robust statistical analysis are anticipated to make FMRI-BCI more widely available and applicable.

Keywords: FMRI (Functional Magnetic Resonance Imaging), BCI (Brain Computer Interface), EEG (Electro Encephalograph), BOLD (Blood Oxygen level dependent).

I. INTRODUCTION

The first BCI was described by Dr. Grey Walter in 1964. Ironically, this was shortly before the first Star Trek episode aired. Dr. Walter connected electrodes directly to the motor areas of a patient's brain. (The patient was undergoing surgery for other reasons.) The patient was asked to press a button to advance a slide projector while Dr. Walter recorded the relevant brain activity. Then, Dr. Walter connected the system to the slide projector so that the slide projector advanced whenever the patient's brain activity indicated that he wanted to press the button. Interestingly, Dr. Walter found that he had to introduce a delay from the detection of the brain activity until the slide projector advanced because the slide projector would otherwise advance before the patient pressed the button! Control before the actual movement happens, that is, control without movement – the first BCI! Unfortunately, Dr. Walter did not publish this major breakthrough. He only presented a talk about it to a group called the Ostler Society in London [1].Today many research is going on in this field, where FMRI is the one technology which is associated with the BCI in this paper.
BCI also called as MCI (Mina- machine interface) is a direct communication pathway between an enhanced or wired brain and an external device. Based on the functional neuroimaging recordings obtained through FMRI, we can easily observe the effects on behavior. In the rest of the paper, we will compare the MRI with FMRI technologies and the working methodology of FMRI with respect to BCI. The development of FMRI in the 1990s, generally credited to Seiji Ogawa and Ken Kwong, is the latest in long line of innovations, including positron emission tomography (PET) and near infrared spectroscopy (NIRS), which use blood flow and oxygen metabolism to infer brain activity [3].Functional magnetic resonance imaging, or FMRI, is a technique for measuring brain activity. It works by detecting the changes in blood oxygenation and flow that occur in response to neural activity – when a brain area is more active it consumes more oxygen and to meet this increased demand blood flow increases to the active area. FMRI can be used to produce activation maps showing which parts of the brain are involved in a particular mental process.
The fact that the magnetic state of hemoglobin changes with its state of oxygenation was discovered in 1936 by Pauling and Coryell, before the discovery of nuclear magnetic resonance (NMR) itself [2].

II. MRI VS. FMRI

Magnetic resonance imaging, or MRI, is a machine used for brain structure imaging. MRI, or magnetic resonance imaging, is the technique widely used in modern medicine to diagnose soft tissues of the human body. The most important application of MRI is in the investigation of the human brain.
FMRI, or functional magnetic resonance imaging, is the technique which utilizes magnetic fields to create the image of what is going inside the human brain. The focus of FMRI is on the processes in the human brain as they will develop in time. FMRI functions through blood flow or blood oxygen level measurements to achieve the brain's functional image. It is primarily used to gather relevant data as to the consumption of oxygen by the tissues. Through its modernization, fMRI sequences will view a picture of the brain's active region by picking up the excess blood supply called Blood Oxygen Level Dependence (BOLD) [4].

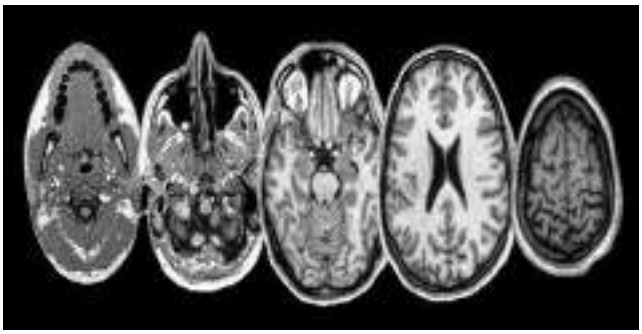


Fig 1: Image obtained through MRI

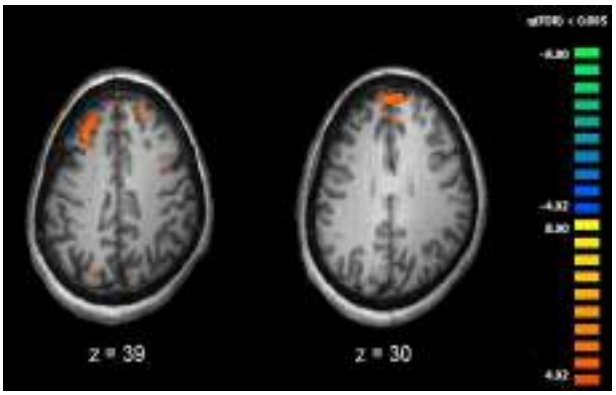


Fig2: Image through FMRI

A high spatial resolution MRI image is used to map the changes in oxygenated and deoxygenated blood ratio which is measured by a series of low spatial resolution fMRI scans, although scans can be adjusted to incorporated higher temporal resolutions at the cost of a lower spatial resolution

FMRI has advantages of other methods of neurological activity as it does not require radioactive tracers to be injected. Finally MRI is both a structural and functional imaging technique, it can provide a range of differently contrasted high spatial resolution images, or alternatively it can provide a series of low spatial resolution at a relatively high temporal resolution to map neurological functionality due to the levels of oxygen within the blood[5].

III. ARCHITECTURE OF FMRI-BCI

An FMRI-BCI system shown in Fig. 1 is a closed-loop system with the following major subsystems: signal acquisition, signal preprocessing, signal analysis, and signal feedback [6].

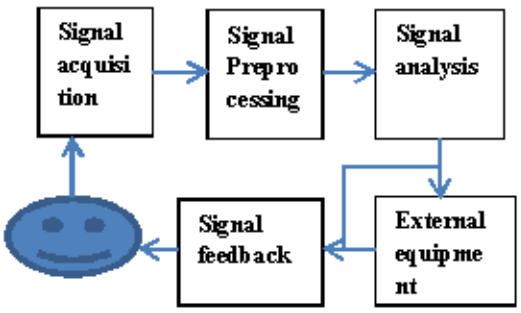


Fig3: Architecture of FMRI-BCI

An echo planar imagining is used to acquire whole brain images from experimental subjects, then the three dimensional brain is divided into a specified number of two dimensional slices. More slices gives better resolution. For the real time achievement of FMRI-BCI requires to acquire once the images have been produced for repetition. Pattern based methods can be used to identify the patterns of activation in the images and then these patterns are used to control the feedback. After specific pattern to be identified is used as a feedback, further processing is conducted to produce a brain activity in the form of functional maps, curves, waves,etc.,

IV. WORKING METHODOLOGY

A human head is equipped with electrocap which helps in the detection of signals of brain. The BCI works in the following steps.

1. Signal production

For a BCI, brain signals need to be produced by human and this can be generated by placing an eletrocap on his/her head. Once the signal is generated from the brain, he will get control over the stimuli.

2. Signal detection

Many signals will be generated from the brain, but most well known are: EEG and FMRI[4]. EEG measures the electrical activity of the brain and FMRI measures the blood-flow in the brain.

3. Signal processing

One of the problem we find in dealing with brain waves is that, it contains lots of noise. This noise needs to be filtered out. The data can now be used for detecting actual signals. The signals which are removed from noise will help for better achievement.

4. Signal transduction

When we have detected the interesting signals, we can use that in interfacing. In our paper, we are using the signals of EEG and FMRI to communicate with the external devices or computer.

EEG records electrical patterns of the brain. The billions of nerve cells in the brain produce very small electrical signals that form patterns called brainwaves. The electrodes in the electro cap detect brain waves and EEG machine amplifies the signals and records them in a wave pattern on graph paper or a computer screens. The following figure shows the capture of EEG wave patterns.

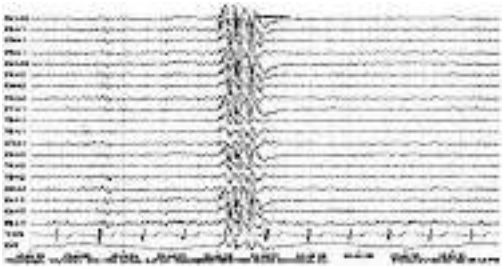


Fig4: A sample EEG recording showing a focal spike typical of a seizure.

FMRI measures the oxygen, which is delivered to neurons by hemoglobin in capillary red blood cells. When neuronal activity increases, there is an increased demand for oxygen and that region of neural activity. Due to this blood oxygenation variation of neuron activity, we can use it to detect the brain activity. FMRI mainly captures the blood- oxygen level dependent (BOLD) regions. The following figure shows the capture of BOLD regions from fMRI.

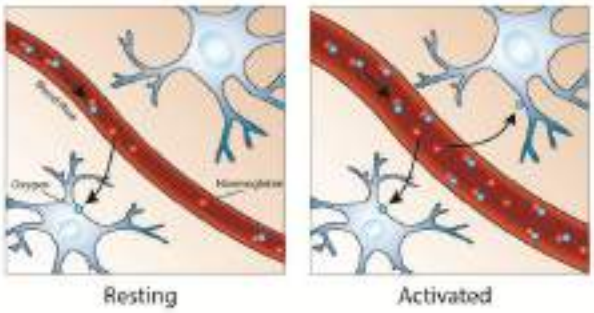


Fig5: FMRI measures the oxygen level in the regions of the brain.

V. CONCLUSION

Brain computer interface based on FMRI enable real – time feedback of brain regions to learn volitional regulation of those regions. FMRI-BCI can be widely applicable due to usage of advanced MRI scanners and faster feedback technologies. Since FMRI identifies the BOLD responses, which will tell us the activity to be performed by the brain, and that can be used to communicate with the external devices by blind people specially.

VI. REFERENCES

[1] D.C Dennett, consciousness explained, Back BAY Books, Lippincoh Willianes and Wilkins,(1192).

[2] Pauling L coryellCD: the magnetic properties and structure of hemoglobin oxyhemoglobin and carbomonoy hemoglobin. Proc natl acadsci USA 22:210-1936

[3] www.psychcentral.com

[4] Mark.F.bera,barry.W. conrros,michaelA. Pasadiso,neuroscience: exploring the brain 3rd edition, wolters klewer health,2007,p178

[5] www.actforlibraries.org

[6] Da-Huan Li,Qin Gao,Wei-Shuai Cu and Hua-Fu chen:FMRI-BCI:a review:journal of Electroni science and techonology of China,vol 7,no 1, march 2009

[7] Principles of BCI based on real time functional magnetic resonance imaging (FMRI); nikdaus weiqkopf, klains mathiak; IEEE transactions on biomedical engineering , vol51,no6,june 2004

[8] Classification of EEG signals in a Brain computer interface system: erill Andreas Larsen, Nor wegian university of science and techonology

[9] Functional MRI: methods and applications: Stuart , the university of Notlinghan,1997

Multi Model Image Registration and Fusion using Fast Discrete Contourlet Transform

YASHASWINI V A, P. BHUVANESWARI, B.ARAVIND
Student, Dept.of Electronics and Communication, RRCE, Bengaluru, Karnataka,India
Associate Professor, Dept.of Electronics and communication,RRCE, Bengaluru,Karnataka,India
Scientist, Space Craft Operation Area, ISTRAC, Bengaluru, Karnataka,India
yashuak@gmail.com, bhuvanasamuel@gmail.com, aravind.rec@gmail.com

ABSTRACT: This paper proposes Transform Domain Fusion Rule (TDFR) via high pass modulation using Local Magnitude Ratio (LMR) in Fast Discrete Contour let Transform (FDCT) domain. Contour let transform method uses low resolution multispectral image and Cartosat-1 Panchromatic (PAN) of spatial resolution 2.5m is used as high resolution panchromatic image. This both images are up sampled in order to resize the image and registered using mutual information to parameterize and solve the correspondence problem in feature-based registration. This fusion rule generates as high resolution multispectral image at 2.5m spatial resolution with the help of Inverse Fast Discrete Contour let Transform increases the contrast of an image and thus gives sharper visual appearance. Different techniques are compared with this method such as Wavelet, Principal component analysis (PCA) and Curve let also tabulated.

Keywords—Fast Discrete Contour let Transform (FDCT), Local Magnitude Ratio (LMR), Panchromatic (PAN), Principal Component Analysis (PCA),Transform Domain Fusion Rule (TDFR).

I INTRODUCTION

The image should focus everywhere to obtain more information, instead of focusing on just one object. Remote sensing image fusion aims at integrating the information conveyed by data, acquired with different spatial and spectral resolutions, for purposes of photo analysis, feature extraction, modeling, and classification.Special domain term refers to more no of pixels combined for composing an image and transform domain enhances the image computing 2-dimensional unitary transform of image [1, 2]. Principal Component Analysis (PCA) fusion will enhance the special quality but represents uniformity features due to dense non zero entries [3].Whereas contourlet Image fusion methods based on injecting high frequency components taken from the panchromatic image into resampled versions of the multispectral data have demonstrated a superior capability of translating the spectral information of the coarse scale multispectral data to the finer scale of the panchromatic image with minimal introduction of spectral distortions.

Image fusion requires the definition of a model establishing how the missing high pass information to be injected into the re-sampled multispectral bands is extracted from the panchromatic image. Firstly, we are considering two input images one low resolution multispectral image i.e. reference image and another one panchromatic image i.e. target image must be pre-processed, the pre-processed images are subjected to Fast Discrete ContourletTransform (FDCT), which converts the time domain into frequency domain images, these frequency domain values are then fuse together by applying Transform Domain Fusion Rule, then Inverse Fast Discrete Contourlet Transform (IFDCT) reconverts the frequency domain values in to time domain values. By all these methodologies we can achieve multi model image registration and fusion using Fast Discrete Contourlet Transforms (FDCT).

Especially image acquired by satellites is helpful in tracking of earth resources; geological mapping; forecast of agricultural crops, metropolitan growth, and climate; flood and free control; and several environmental applications. Image broadcast and storage applications happen in broadcast television, teleconferencing, broadcast of facsimile images for office computerization communication through computer networks, closed-circuit television based safety monitoring systems, and in armed forces communications.

I. RELATED WORK

Chen Chen et.al [4] has proposed thatimages of low resolution multispectral image and high resolution panchromatic imagecan be simultaneously fused with in the geographical area based on fast discrete curvelet transform. Earlier spectral information of the multispectral images where preserved by virtue of the dynamic gradient sparsity property. The linear computational complexity in the size of the output image in each iteration is solved by the efficient algorithm. High-quality products are obtained from coarsely registered real-world IKONOS data sets from four satellite images. Experimentation based results have shown appropriate improvements in Entropy, Peak signal to noise ratio (PSNR) and Root Mean Square Error (RMSE). C.V.Rao et.al [5]has proposed satellite image fusion based on fast discrete curvelet transform, which retains both special and spectral qualities of source image.This curvelet transform overcomes the drawbacks of convention 2-dimensional discrete wavelet transform. Each multispectral band is compared to the high resolutionpanchromatic image for special quality evaluation and resampled multi spectral bands of Linear Imaging Self Scanner (LISS-4) sensor image and corresponding bands in the fused image are compared for spectral quality evaluation.

II. Problem Statement

High resolution satellite imagery provides researchers with information sources necessary for use in many change detection applications. However, problems occur when applying conventional traditional image processing methods to process and analyze these high resolution images. A common problem of existing techniques is the color distortion that occurs in the resulting fused images. For high resolution images, such as IKONOS and QuickBird, the wavelength of the panchromatic image is much broader than that of the multispectral bands. This discrepancy between the wavelengths causes considerable color distortion to occur when fusing high resolution panchromatic and multispectral images. To solve the color distortion problem, methods based on wavelet have been introduced and have demonstrated superior performance. However, when applying wavelet based methods, spatial distortions, typically ringing or aliasing effects, and originating shifts or blur of contours and textures may occur.Wavelet basis is isotropic and takes lot of co-efficient to account for edges [8-10]. These problems- which can be as pronounced as color distortion mentioned above - are emphasized by misregistration between a panchromatic image and its multispectral counterparts, especially if the wavelet is not shift-invariant.

III. Proposed System

Fig.1 shows the block diagram of proposed system, the inputs low resolution multispectral image and high resolution panchromatic image are up-sampled. Here low resolution multispectral images are RGB color image and high resolution PAN image are grey scale image. The up-sampled color bands and the panchromatic band are aligned to reduce artifacts due to Mis-registration. There are different methods used in image registration that enhances the visual appearance of an image [6, 7].The R, G, B color bands are up-sampled to the same resolution as the panchromatic band. Panchromatic band is then directly substituted to Fast Discrete Contourlet Transform which provides a flexible multi-resolution, local and directional expansion of images. The multi resolution images are registered using mutual information which is used to have identical geometry and size in feature-based registration. The obtained registered image is merged using Transform Domain Fusion Rule. Inverse Fast Discrete Contourlet Transform increases the visual appearance so that images are more clearly depicted. Finally images are fused to obtain high resolution multi-spectral image.

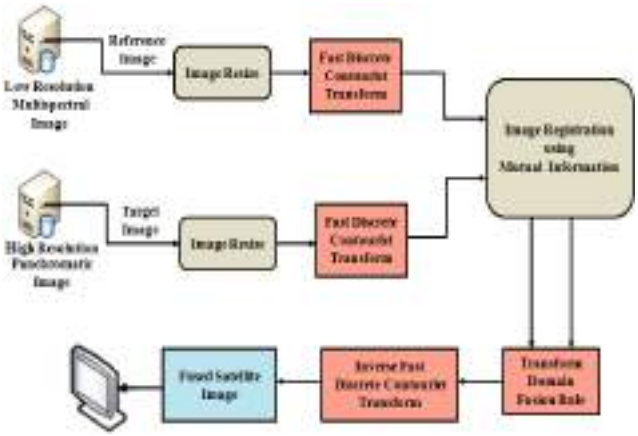


Fig.1 Block Diagram of Proposed Work

4.1 Fast Discrete Contourlet Transform (FDCT)

When compare to other existing transforms theFast Discrete Contourlet Transform (FDCT) are simpler, faster, and very less redundant, with the same complexity of rapid inversion algorithms the Fast Discrete Contourlet Transform (FDCT) are invertible. Coutourlet is a method of two-dimensional image representation. There are two-layer filters in contourlet transform they are, Laplacian pyramid transform is used to realize multi-scale decomposition and multi direction decomposition can be implemented by directional filter bank.Laplacian pyramid technique is used for data compression and in directional filter bank, frequencies can be coded to finer resolution after the decomposition.The two filters of the countourlet transform are used for image decomposition and yield a smoother sparse representation of the original image [11-13].Contourlet transform can realize multi-resolution and multi-direction decomposition flexibly and grasp essential character of image.The important characteristics of contourlet transform are directionality, multi-resolution, localization, anisotropy and critical sampling [14,15].Contourlet transform presented by digital conversion of discrete domain first, and then extended to continuous domain and analysis are done on the basis of properties when it is compared with curvelettransform.The curvelet de-noising scheme is very effective because the energy of the signal is captured within few energy transform values but Contourlet transform accuracy is more accurate compared to curvelet [16].

To present the simulation, contourlet are more convenient. Fast Discrete Contourlet Transform (FDCT) provides variable frequency components locally for synthesis and analysis of digital image in multi-resolution analysis. The main aim to use Fast Discrete Contourlet Transform is to provide a flexible multi-resolution, local and directional expansion of images. Decomposition sampled signals of fast and structured curvelet gives the anisotropy scaling relation for curves by applying contourlet transforms. As an output, the proposed contourlet transform provides an incompact representation for two-dimensional piecewise smooth signals that compares images. The link between the developed filter

banks and the continuous-space constructions is set up precisely in a newly defined directional multiresolution analysis. Finally, we show some numerical experiments demonstrating the potential of the new transform in several image processing tasks.

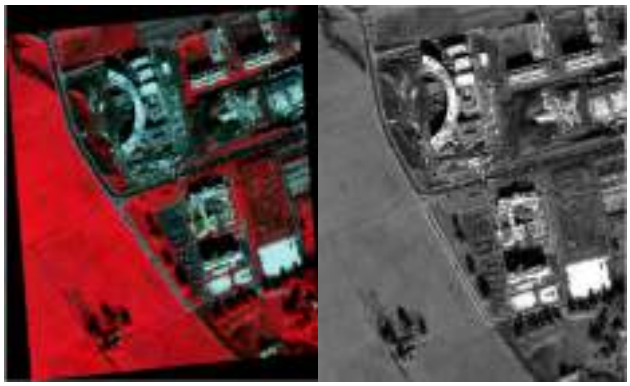
Fast Discrete Contourlet Transform (FDCT) is multi-scale geometric transform, which is a multi-scale pyramid with many directions and positions at each length scale. Fast Discrete Contourlet Transform (FDCT) has main direction associated with 2-dimensional anisotropic extension to classical wavelet transform. Fast Discrete Contourlet Transform (FDCT) can be translated and dilated analogous to wavelet. Whereas curvelet are based on multiscale ridgelets that separates images into disjoint scales with the help of bandpass filtering [17-20]. The frequency content of the Contourlet is controlled by the dilated scale index; direction can be changed through a rotation along with indexed position. This rotation is indexed by an angular index. Contourlet satisfy anisotropic scaling relation, which is generally referred as parabolic scaling.

Fast Discrete Contourlet Transform (FDCT) is applied to a rotated and un-sampled high-resolution grid. At Contourlet domain the high resolution grid is decomposed at three levels. The locations must be determined in each sub-band in order to interpolate the missing pixels. In Contourlet domain missing pixels corresponds to missing coefficients of each sub-band. The missing coefficients are interpolated at finest scale. Inverse Contourlet transform reconstructs the original high-resolution grid.

4.2 Transform Domain Fusion Rule

In transform domain fusion rule firstly, both low resolution multispectral image and panchromatic images are pre-processed and both images must be at identical geometry and of same sizes, then extract band wise the multispectral data in Green, Red and near infra-red bands, after getting multispectral data, apply Fast Discrete Contourlet Transform (FDCT) to both low resolution multispectral image and panchromatic image. Input images are decomposed into four levels in multiple directions. At lower frequencies multispectral image and the fused image fusion rule 2 is explained and applied. Fusion rule 1 is defined for the multidirectional multi resolution Contourlet co-efficient at higher frequencies based on high pass modulation. A set of Contourlet planes are constructed for the fused image and Inverse Fast Discrete Contourlet Transform (IFDCT) is applied to reconvert resolute frequencies in to time domain values. Combination of three resultant fused bands provide multispectral fused image.

IV. Experimental Result



(a) (b)

Fig.2 Input Images

(a) Low Resolution Multispectral Image (Reference Image)
(b) High Resolution Panchromatic Image (Target Image)

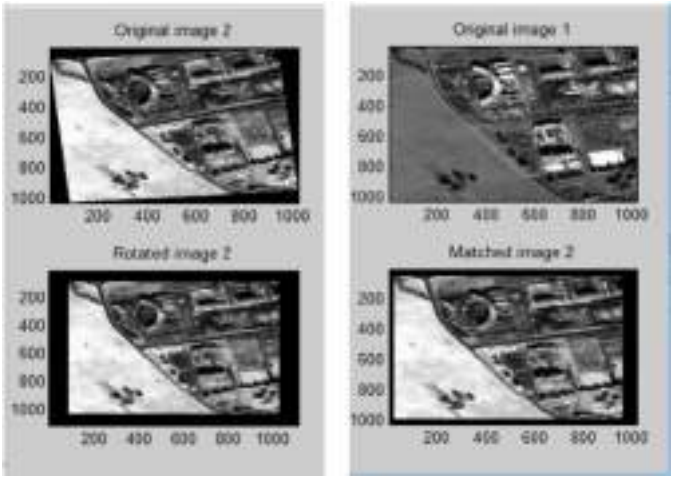


Fig.3 Level 1 Resizing of an Image into RGB plane

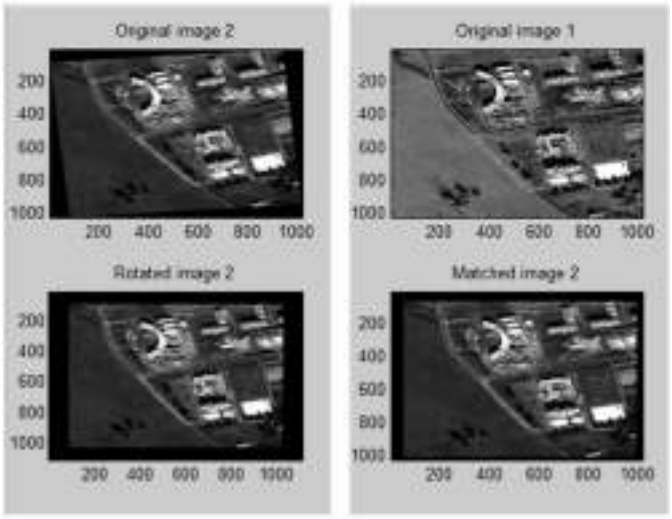


Fig.4 Level 2 Resizing of an Image into RGB plane

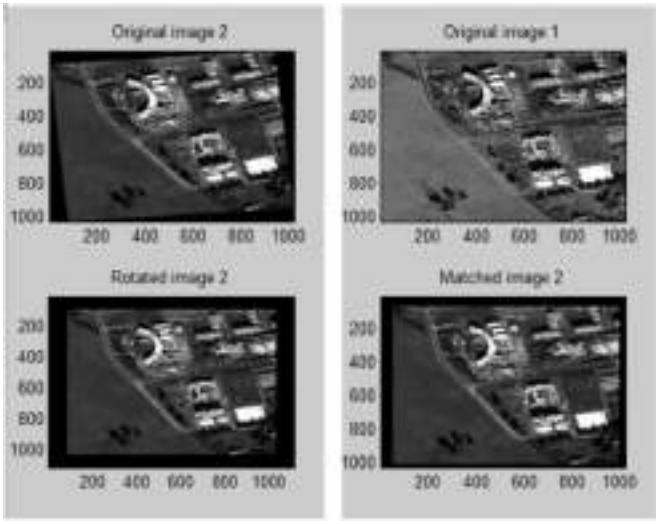


Fig.5 Level 3 Resizing of an Image into RGB plane

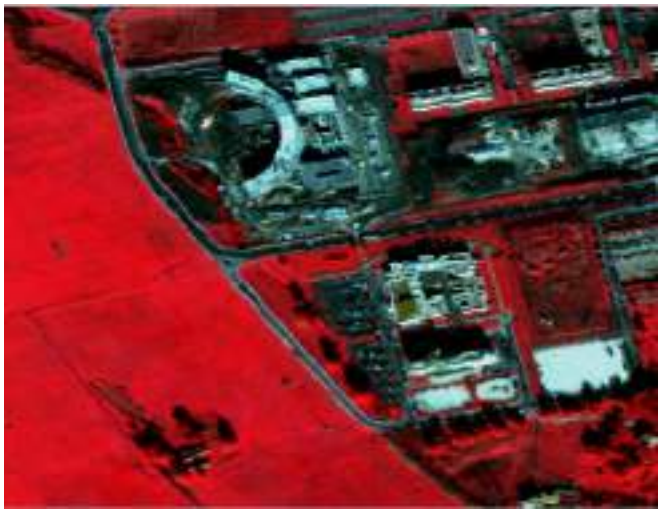


Fig.6 Registered Image of Reference and Target Image

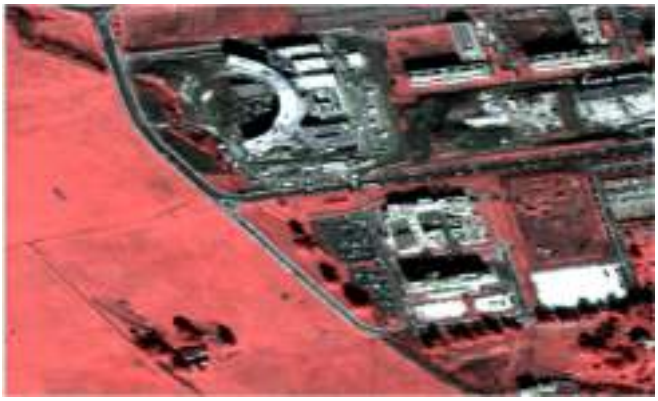


Fig.7 Finally Fused Image of High Resolution Multispectral Image.

V. PERFORMANCE EVALUTION

Table 1.Parameters value comparison of various methods

METHOD	PARAMETERS			
	RMSE	PSNR	ENTROPY	STANDARD DEVIATION
PCA	24.1	20.7	2.8944	47.5389
Wavelet	4.5	36.1	5.1801	53.4367
SIRF	4.1	47.5	6.2147	59.9738
Contourlet	3.5	49.3	7.6261	62.9543

VI. COMPARISON PLOT

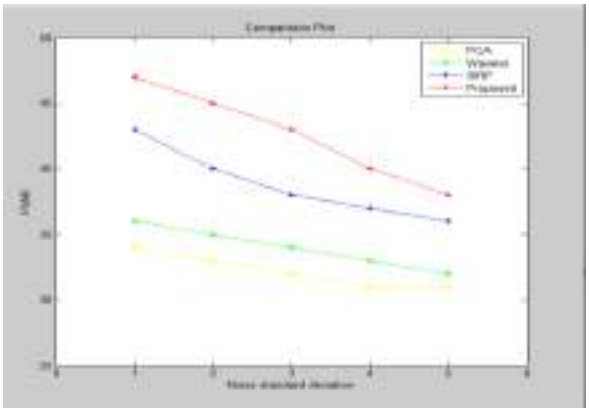


Fig.8 The PSNR values of PCA,Wavelet,SIRF and contourlet with different level of noise.

VII. CONCLUSION

High resolution satellite imagery provides researchers with information sources necessary for use in many change detection applications. However, problems occur when applying conventional traditional image processing methods to process and analyze these high resolution images. A common problem of existing techniques is the color distortion that occurs in the resulting fused images. For high resolution images, such as IKONOS and QuickBird, the wavelength of the panchromatic image is much broader than that of the multispectral bands. This discrepancy between the wavelengths causes considerable color distortion to occur when fusing high resolution panchromatic and multispectral images. To solve the color distortion problem, methods based on wavelet have been introduced and have demonstrated superior performance. From this proposed system, input low resolution multispectral image and PAN image resized and registered using mutual information before fusing the images, after image registration image is fused using fast discrete Contourlet transform and inverse transform domain based fusion rule is used to fuse the images.

REFERENCES

[1] "Image Enhancement using a Combined Approach of

Spatial and Transformation Domain Techniques",SnehalO.Mundhada, Prof. V. K. Shandilya,International Journal of Emerging Research in Management &Technology, Dec 2012.

[2] "Comparative Study of Image Fusion Techniques based on Spatial and Transform Domain", Sweta K. Shah, Prof. D.U. Shah, International Journal of Innovative Research in Science, Engineering and Technology, Vol 3, Issue 3, March 2014.

[3] "Fast and Efficient Medical Image Compression Using Contourlet Transform (FEMI-CCT)", Sandeep Kumar, NitinGoel, Vedpal Singh, AmitChaudhary, NirmalSirohi, Gurbaj Singh,Open Journal of Computer Sciences,Vol 1, May 2013.

[4] "SIRF: Simultaneous Satellite Image Registration and Fusion in a Unified Framework",Chen Chen, Yeqing Li, Wei Liu, and Junzhou Huang, IEEE Transactions on Image Processing, Vol. 24, No. 11, November 2015.

[5] "Satellite Image Fusion using Fast Discrete Curvelet Transforms", C.V.Rao, J.MalleswaraRao, A.Senthil Kumar, D.S.Jain, V.K.Dadhwal, National Remote Sensing Centre, Indian Space Research Organization, Hyderabad-500037, India, IEEE 2014.

[6] "Image Registration Techniques: An overview", Medha V. Wyawahare, Dr. Pradeep M. Patil, and Hemant K. Abhyankar,International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No.3, September 2009.

[7] "Image Registration and Wavelet Based Hybrid Image Fusion", Mrs.Disha S Bhosle Mrs. Kanchan S Gorde, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 2, and Ver. I (Mar-Apr. 2014).

[8] "Image Fusion based on Wavelet and Curvelet Transform using ANFIS Algorithm", NavneetKaur and MadhuBahl, International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 9, September 2014.

[9] "A New Image Fusion Method based on Integration of Wavelet and Fast Discrete Curvelet Transform", Anjali A. Pure, Neelesh Gupta and MehaShrivastava, Vol 69, Issue 18, May 2013.

[10] "A Novel Architecture for Wavelet based Image Fusion", SusmithaVekkot, PanchamShukla, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol 3, No 9, 2009.

[11] "An improved contourlet transform technique for image watermarking", Y.Anusha, K.Veeraswamy, International Journal of Research in Computer andCommunication Technology, Vol 2, Issue 11, November- 2013.

[12] "Gray Scale Image Fusion using Modified Contourlet Transform",Gurdeep Singh, Naveen Goyal,International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 9, September 2013.

[13] "Image Fusion Using Improved Contourlet Transform Technique", Nupur Singh, PinkyTanwar, International Journal of Recent Technology and Engineering (IJRTE), Vol 1, Issue 2, June 2012.

[14] "Fast and Efficient Medical Image Compression Using Contourlet Transform (FEMI-CCT)", Sandeep Kumar, NitinGoel, Vedpal Singh, AmitChaudhary, NirmalSirohi, Gurbaj Singh,Open Journal of Computer Sciences,Vol 1, May 2013

[15] "Fuzzy Based Image Compression on ROI using Optimized Directional Contourlet Transform",M.Tamilarasi , Dr.V.Palanisamy,International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2009.

[16] "Image Denoising using Curvelet Transform", AmandeepKaurBains, PrabhneetSandhu, International Journal of Current Engineering and Technology, Vol.5, No.1, Feb 2015.

[17] "Remote Sensing Satellite Image Fusion Using Fast Curvelet Transforms",Namratha H. N, Raghu M. T, International Journal of Science and Research (IJSR), Vol 4 Issue 4, April 2015.

[18] "Image Compression Algorithm Using a Fast Curvelet Transform", Walaa M. Abd-Elhafiez, International Journal of Computer Science and Telecommunications, Vol 3, Issue 4, April 2012.

[19] "An Image Fusion using Wavelet and Curvelet Transforms", Smt.G. Mamatha, L.Gayatri, Andhra Pradesh, India, Global Journal of Advanced Engineering Technologies, Vol1, Issue-2, 2012.

[20] "Spatial Images Feature Extraction Based on Bayesian Nonlocal Means Filter and Improved Contourlet Transform", Pengcheng Han, Junping Du,Hindawi Publishing Corporation, Journal of Applied Mathematics, Volume 2012.

An Effective & Automated MR Brain Image Segmentation & Tumor Detection

Pushpa Latha N,Aruna Kumara B
PG Scholar, Dept. of CSE, RRCE, Bengaluru
Assistant Professor, Dept. of CSE, RRCE, Bengaluru
pushpa.ln18@gmail.com,arun.bethur87@gmail.com

ABSTRACT: The image processing is an interesting and challenging field now a days and medical image processing plays a major role in it. The medical images are used to analysis the diseases like brain tumor, cancer, diabetes, etc. The brain tumor is one of the dangerous diseases where many people suffer from this disease. Image segmentation is used to take out the suspicious parts from medical images like MRI, CT scan, and Mammography etc. For MRI brain image segmentation adaptive k means clustering is used. The feature extraction is done using the GLCM (Gray Level Co-occurrence Matrix) which stay away from the creation of misclustered region. The feature selection is done to improve the classifier accuracy using PCA (Principle Component Analysis). A PSVM (Proximal Support Vector Machines) classifier is used to automatically detect the tumor from MR brain image which is faster and efficient than the existing method SVM.
Keywords - Feature Extraction, Image Segmentation, MRI, Proximal Support Vector Machines

I. INTRODUCTION

The Magnetic Resonance Imaging (MRI) is usually used medical image technique which provides detailed information of the internal tissue constitutions of the image. They provide complete information about brain tumor anatomy, cellular structure and vascular supply. It provides an effective result for the diagnosis of disease provides efficient treatment and monitors the disease in an efficient manner.

The analysis of brain tumor can be examined by MRI scan or CT scan. The main advantage of MRI scan is that it does not contain any radiations, whereas CT scan contains many radiations. So that MRI does not affect the human body. MRI is better compared to CT scan. MRI is a type of scanning device, which use magnetic field and radio waves. MRI scans provide exact picture of anatomical structure of tissues. A tumor may be primary or secondary. It's said to be primary if the tumor is at the beginning state. It's said to be secondary if the tumor has spread to other places and has grown.

The brain tumor affects CSF (Cerebral Spinal Fluid) and causes strokes. The treatment is given for the strokes rather than the treatment for tumors. So the detection of the tumor is important for that treatment. If the tumor is detected at an earlier stage the life time of the person will be increased. In image processing an array of pixels is taken as input, which produces another array of pixels as output which represents perfection to the original array. The Digital Image Processing processes the two-dimensional picture by a digital computer changing the existing image in the required manner. This process removes the noise, improves the contrast of the image, and removes blurring. If an image is low contrast and dark, the images improve its contrast and brightness. The histogram equalization improves all parts of the image when the original image is irregularly illuminated. The images contrast is increased using the enhancement techniques and also the quality of the images are improved using the image enhancement techniques.

II. RELATED WORK

The process of collection of content that provides a wide area view of the project to be taken up is Literature Survey. For thorough study of the system it needs to go through each and every technical aspect of the related material.

A. MRI over CT scan

MRI is used because it's provide accurate visualize of anatomical structure of tissues. The analysis of brain tumor can be found by MRI scan or CT scan. The main benefit of MRI over CT scan is, MRI does not contain any radiation. It provides accurate picture of anatomical structure of tissues. It does not affect the human body. Fundamentally MRI is better compared to CT scan, which uses magnetic field and radio waves.

B. K-Means Clustering and Fuzzy C-Means Algorithm

Tumor is an uncontrolled growth of tissue in part of the body. There are different types of algorithms developed for brain tumor detection. But there are some drawbacks in the detection and extraction. The segmentation process is done through k-means clustering and fuzzy c-means algorithms. After the segmentation process the brain tumor is detected and its exact location is identified. When comparing to the other algorithms the performance of fuzzy c-means plays the most important role. The patient's stage is identified by this process and it's checked whether it can be cured with medicine or not.

C. Image Enhancement

Image enhancement is used for the purpose of image processing. To improve the digital image quality image enhancement is used. Image histogram is used in image enhancement. The histogram in the framework of image processing is the operation by which the occurrences of each intensity value in the image is shown and histogram equalization is the method by which the dynamic range of the histogram of an image is increased.

D. Image Segmentation with Non Negative Matrix Factorization

Nonnegative Matrix Factorization is one of the techniques to reduce the dimensionality of the data. NMF is used for the image processing methods such as pattern analysis and text mining. It's used as an uninterruptable decomposition approach for detection of tumor and for further classification and for feature extraction. NMF finds two non-negative matrices where the product closely approximates the original matrix. NMF contains all matrices that contain only non-negative elements.

E. MRI Brain Image Classifications

It presents the techniques for the classification of the magnetic resonance images (MRI). There are three stages in the technique, which are feature extraction, dimensionality reduction, and classification. In the first stage, the features related to MRI images are obtained using discrete wavelet transformations (DWT). In the second stage, the most essential features of the magnetic resonance images have been reduced, the reduction is done using principal component analysis (PCA). For classification two classifiers are used. The first classifier is based on the feed forward back propagation artificial neural network (FP-ANN) and the second classifier is based on the k nearest neighbor (k-NN). The classifiers are used to classify the MR images as normal or abnormal. The classification is obtained with a success of 97% using FP-ANN and success of 98% using k-NN. It shows that the technique is robust and effective compared with other works.

III. SYSTEM ARCHITECTURE

The proposed architecture gives an efficient way to detect the tumor with a high accuracy. The architecture has two phases training and testing, in training the MRI brain image is done with pre-processing, segmentation, feature extraction, feature selection and it's stored in the hospital database. Some of the sample images are taken from the database and testing is performed. In testing phase a query image (MR brain image) is taken to find whether it's normal or abnormal. The query image will also undergo pre-processing, segmentation and feature extraction with new technologies. Once feature extraction is done the image is classified using PSVM classification, the classification done through PSVM gives the high accuracy detection compared to SVM. The tumor type is declared in the results.

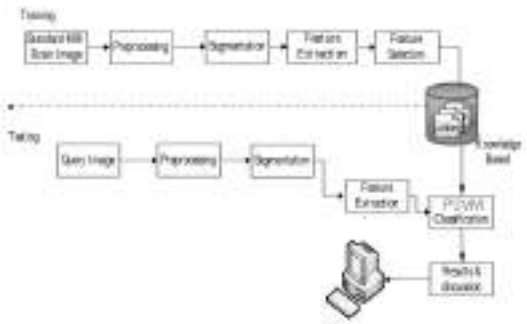


Figure 1: Brain Tumor Diagnosis System

In existing system, the system uses different segmentation methods. They use clustering algorithms like fuzzy c means, k means which has more disadvantage than the proposed system. For segmentation the existing system uses k means algorithm, where it has some disadvantages. To overcome the disadvantages, an adaptive K-means clustering algorithm (AKM) is proposed in this system. In existing for tumor classification the system uses SVM (Support Vector Machines). This is very slow process and complex when comparing the proposed system. The accuracy rate and precision rate is very less when compared to the proposed system.

IV. IMPLEMENTATION

The brain tumor diagnosis system is implemented using matlab. The brain MR image is taken and the following processes are done one by one to detect the tumor type. The processes are:

i. Image Pre-processing
ii. Segmentation
iii. Feature Extraction
iv. Image Classification

Image Pre-processing

Image pre-processing is performed to remove noise and clean up the image background, suppresses unnecessary distortions or enhances some image features important for further processing. Adaptive filters are used for image pre-processing where it can complete some signal processing tasks that traditional digital filters cannot.

Segmentation

Segmentation is the process of dividing an image into regions with similar properties such as contrast, color, brightness, gray level, color, and texture. Gray scale image is obtained from the color image. Adaptive k means clustering algorithm is used for the segmentation process. Tumor regions are effectively segmented by adaptive k means clustering algorithms and thus the tumor portion from MRI image is detected.

Feature Extraction

In feature extraction the texture feature is defined by using Gray Level Co-occurrence Matrix (GLCM). The important features for brain tumor recognition like contrast, homogeneity, energy and entropy are extracted using GLCM. The features extracted gives the property of the texture, and are stored in knowledge base and further compared with the features of unknown sample image for classification. The texture features of the MRI brain images are used to classify the tumors.

Image Classification

For automatic detection of tumors at a faster rate, PSVM (Proximal Support Vector Machines) is used as the classifier for brain image classification using matlab with GLCM.

PCA (Principal Component Analysis) is used for feature selection, which is used for improving the detection accuracy. Using the features extracted with GLCM the PSVM classifier classifies the tumor type.

V. CONCLUSION

The effective segmentation and classification is proposed using adaptive k means and PSVM. After segmentation, feature extraction is performed using GLCM, and then the images are classified using PSVM classification and feature selection using PCA. At the training phase of PSVM, the texture features are utilized which can reduce the computation complexity of PSVM classifier. The result shows that the proposed system shows a high accuracy rate and less error rate. In future the system can be improved to support other types of cancer images. It is necessary to support large number of input and should improve the accuracy rate. To achieve this more number of features can be added with the utilization swarm based feature selection to improve the tumor detection and the classification result.

REFERENCES

[1] Jay Patel and Kaushal Doshi," A Study of Segmentation Methods for Detection of Tumor in Brain MRI", *Advance in Electronic and Electric Engineering. ISSN 2231-1297, Volume 4, Number 3* (2014), pp. 279-284.

[2] Alan Jose, S.Ravi, M.Sambath ," Brain Tumor Segmentation Using K-Means Clustering And Fuzzy C-Means Algorithms And Its Area Calculation", *International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization)Vol. 2, Issue 3,* March 2014.

[3] Sapana S. Bagade, Vijaya K. Shandilya, "Use Of Histogram Equalization In Image Processing For Image Enhancement", *International Journal of Software Engineering Research & Practices Vol.1,* Issue 2, April, 2011

[4] J. Alirezaie, M. Jernigan, C. Nahmias, Automatic segmentation of cerebral MR images using artificial neural networks, *IEEE Transactions on Nuclear Science* 45 (4) (1998) 2174–2182.

[5] T. Logeswari, M. Karnan, Hybrid self-organizing map for improved implementation of brain MRI segmentation, in: *International Conference on Signal Acquisition and Processing,* 2010.

Classification of Kidney Stone Using GLCM

Dr.Punal M Arabi, Surekha Nigudgi, Rohith N Reddy, Dhatri
Department of Biomedical Engineering, ACSCE, Bangalore-74
arabi.punal@gmail.com, sursanju@gmail.com, rohithnreddy12345@gmail.com, monanatraj10@gmail.com

ABSTRACT : Kidney stone disease is one of the risks for life throughout the world and majority of people with stone formation in kidney at the initial stage do not notice it as disease and it damages the organ slowly. Current estimation is that there are 30 million people suffering by this disease. The currently available widely used low cost imaging techniques for diagnosing kidney diseases include X-ray imaging and Ultrasound imaging. This paper proposes an approach for the classification of different types of Kidney stones using gray level co-occurrence matrix (GLCM) features namely contrast, co-relation, energy and homogeneity. Different types of Kidney stones namely Struvite stones, Stag horn stones and Renal Calculi stones were analyzed and the results obtained show that the proposed GLCM feature analysis method could be used to classify kidney stones. The results obtained further show that there is a possibility of developing CAD and computer aided classification of kidney stones by the proposed texture analysis method.

Keywords – Classification, GLCM, Renal Calculi, Stag horn, Struvite, texture analysis.

I Introduction

Kidney diseases are on rise throughout the world and majority people with kidney disease do not notice the disease as it damages the organ slowly before showing symptoms. The increasing number of patients with kidney diseases leads to a high demand of early detection and prevention of kidney diseases. It is well known that ultrasound (US) can be used as an initial evaluation to estimate kidney size and position, and help to diagnose structural abnormalities as well as presence of cysts and stones. However, diagnosis of kidney diseases and abnormalities using ultrasound demands decision from experts as US images suffer from speckle noise. Speckle has variation of gray level intensities. Therefore, to enhance quality of these images, some image processing techniques are usually applied for better understanding of hidden information as well as for extracting some parameters or features that will be useful for diagnosis of the images. Current estimates are that 30 million (1 in 11) Americans will experience a kidney stone within their lifetime, and up to 50% of new stone formers will have a recurrence, within as early as 5 years . The data suggest the incidence of kidney stones will continue to grow with our increasing obesity and diabetes rate, and even climate change. When dietary minerals in the urine become supersaturated, crystals of Urinary stones are formed. Stones almost always start in the kidneys which may cause problems or may not be noticed until they move into the urethra (the tube that connects each kidney to the urinary bladder). Once stones pass down the urethra into the bladder, they usually then are passed with the urine, but sometimes they can get into the bladder and grow larger there. The most common symptom of kidney stones is pain in upper back. When the pain is severe there is possibility of getting nausea as well. There can be blood in the urine and also a urinary tract infection. Stones are diagnosed with CT scans, X-rays, or ultrasound. A kidney stone may not cause symptoms until they move within your kidney or until it passes into urethra — the tube connecting the kidney and bladder. Severe pain in the side and back, below the ribs pain may spread to lower abdomen and Pain that comes in waves and fluctuates in intensity, Pain on urination, Cloudy or foul-smelling urine, Nausea and vomiting, Persistent need to urinate, Urinating more often than usual Fever and chills if an infection is present.

II Literature Survey

Farid.G.Mitri et.al proposed vibroacoustography technique for imaging and detecting kidney stones within the kidney and also to show the local anatomical features while differentiating stones from surrounding tissue structures [1]. P.R.Tamiselvi et.al developed a semiautomatic region growing algorithm for ultrasound kidney images to detect calculi from renal calculi images [2]. K.Viswanath et.al proposed kidney stone detection from ultrasound images by level set segmentation and ANN classification [3]. K.Bommanna Raja et.al proposed a fuzzy neural system to offer classification efficiency and to identify the category of kidney stones [4]. K.Dhanalakshmi et.al developed and implemented a computer aided decision support system for an automated diagnosis and classification of ultrasound kidney images [5]. Mahdi Marsousi et.al proposed a new automated kidney detection approach using three dimensional morisons's pouch ultrasound images along with a shape based method to segment detected kidneys [6]. K.Divya Krishna et.al presented a computer aided automatic detection of abnormality in kidney on IOT enabled portable ultrasound systems [7].

III METHODOLOGY

In the proposed method, the acquired images undergo a pre-processing stage which is of two folds namely-

1) Image enhancement for removal of blur.
2) Filtering by low pass filter.

Region of interest is then selected. The selected region of interest is an RGB image which is then converted to gray image.

For this work a set of different types of kidney stone images (three images for each set) are taken for experimentation. The images are analyzed using GLCM parameters.

GLCM ANALYSIS

From every image, the stone is cropped which is the region of interest and the GLCM parameters of the cropped portion are then found.

GLCM matrix has been found using the formulae shown below:

Contrast = $\sum_{(i,j)} [i-j]^2 P(i,j)$ .....(1)

Correlation = $\sum_{(i,j)} P(i,j) \left[\frac{(i-\mu)(i-\mu)}{\sqrt{(\sigma)2(\sigma)2}}\right]$ .....(2)

Energy = $\sum_{(i,j)} P(i,j)^2$ .....(3)

Homogeneity = $\sum_{(i,j)} \frac{P(i,j)2}{1+(i-j)}$ .....(4)

Where p=image

i,j=coordinates

p(i.j)=Intensity value at i,j

The average values of all the parameters are found and using these average values the texture analysis is done.

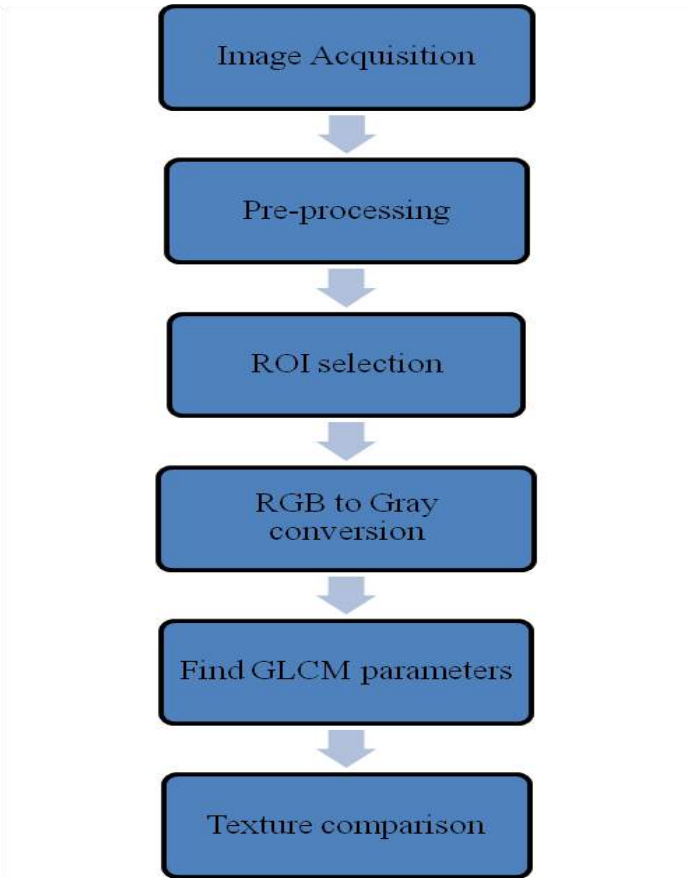


Fig (1) Flow diagram

IV RESULTS

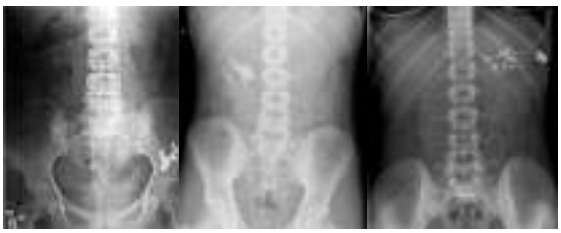


Fig (2) b: Shows the set of Stag horn images

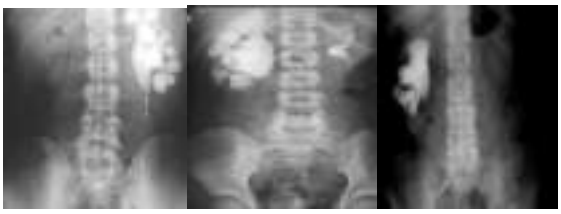


Fig (2) c: Shows the set of Renal Calculi images

Table1. GLCM parameters for Struvite images

Images	Contrast	correlation	energy	homogeneity
1	383.295	0.0165	5.2177	0.1278
2	421.455	-0.0021	4.3463	0.1225
3	444.461	0.0200	5.4641	0.1190
Average	416.403	0.3314	5.0093	0.1231

Table2. GLCM parameters for Stag horn images

Images	contrast	correlation	energy	homogeneity
1	77.5444	-0.0018	0.0022	0.2197
2	89.8888	-0.0013	0.0019	0.2054
3	98.0452	0.0176	0.0019	0.1969
Average	88.4928	0.0048	0.0020	0.2073

Table3. GLCM parameters for Renal Calculi images

Images	contrast	correlation	energy	homogeneity
1	77.5444	-0.0018	0.0022	0.3665
2	89.8888	-0.0013	0.0019	0.2769
3	98.0452	0.0176	0.0019	0.1782
Average	88.4928	0.0048	0.0020	0.2738

Fig (2) a: Shows a set of Struvite images, Fig (2) b: Shows a set of Stag horn images, Fig (3) c: Shows a set of Renal calculi images, Table(1) GLCM parameters for Struvite images, Table(2) GLCM parameters for Stag horn images, Table(3) GLCM parameters for Renal Calculi images.

IV DISCUSSION

A set of three images of Struvite, Stag horn and Renal Calculi Kidney stone X-ray images taken for analysis by the proposed method. These stone images area first pre processed and the regions of interest are then selected. The selected region of interest is an RGB image which is converted to gray image. The average values of GLCM parameters Contrast, correlation, energy, homogeneity are found and tabulated.

The tabulated results show that the Struvite stones have highest contrast, correlation and energy where as its homogeneity is low; Stag horn stones have medium contrast, homogeneity where as its correlation and energy are low. Renal Calculi stones have highest homogeneity where as the contrast is low. Correlation of Renal Calculi stones is of medium value and their energy is though slightly higher than Stag horn stones but much lower in value compared to Struvite stones.

After analyzing about 50 images the average values can be found and from which reference values be fixed. Based on the reference values a scouring system may be developed for computer aided diagnosis and classification of Kidney stones.

V Conclusion

The results obtained show that the texture features could be used to classify kidney stones. The results obtained further show that there is a possibility of developing CAD and computer aided classification of kidney stones by texture analysis method and framing a suitable decision rule. By analyzing many more images by GLCM and other statistical methods a suitable decision rule can be found in future.

ACKNOWLEDGEMENT

The authors thank the Management and Principal of ACS College of Engineering, Mysore road, Bangalore for permitting and supporting to carry out the research work.

REFERENCES

[1] Farid G. Mitri and Randall R. Kinnick "Vibroacoustography Imaging of Kidney Stones *In Vitro"* IEEE transactions on biomedical engineering, vol. 59, no. 1, january 2012

[2] Dr P R Tamiselvi "Detection of Renal Calculi Using Semi Automatic Segmentation Approach" International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 3, May 2013.

[3] K.Viswanath and R. Gunasundari "Design and analysis performance of Kidney Stone Detection from Ultrasound Image by Level Set Segmentation and ANN Classification "2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI).

[4] K.Bommanna Raja and M.Madheswaran", Determination of Kidney Area Independent Unconstrained Features for Automated Diagnosis and Classification" International Conference on Intelligent and Advanced Systems 2007.

[5] K. Dhanalakshmi, Dr. V. Rajamani "An Efficient Association Rule-Based Method For
Diagnosing Ultrasound Kidney Images" 978-1-4244-5967-4/10/$26.00 ©2010 IEEE

[6] Mahdi Marsousi, Konstantinos N. Plataniotis and Stergios Stergiopoulos "Shape-Based Kidney Detection and Segmentation in Three-Dimensional Abdominal Ultrasound Images" 978-1-4244-7929-0/14/$26.00 ©2014 IEEE.

[7] K. Divya Krishna, Vivek Akkala, R. Bharath, P. Rajalakshmi, Abdul Mateen Mohammed "FPGA based Preliminary CAD for Kidney on IoT Enabled Portable Ultrasound Imaging System" 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom).

Authenticating Login session using Mouse biometric with AES Encryption

Shruthi G., Shilpa S.G., Geeta S. Hukkerimath
Department of Computer Science, VTU, Bangalore
shruthigmysore@gmail.com, shilpasg.06@gmail.com,blh_geeta@gmail.com

ABSTRACT- The mouse biometric is a behavioural biometric technology that extracts and analyzes the movement characteristics of the mouse input device when a computer user interacts with a graphical user interface for identification purposes. In this paper, we present a new method for Authenticating login session Using Mouse Biometrics. Mouse dynamics are defined by the characteristics which are acquired by analyzing the inputs a user performs by a pointing device. These characteristics are then combined to form the *factors* which are stored in the user profile. Factors are encrypted using AES Algorithm. During Login session, comparing the current input stream with the previously stored user profile, mouse dynamics can detect anomalies in the user's behavior and stop intrusions while they are happening.The paper proposes mouse based authentication, the application of behavioural biometrics for authentication as a safeguard against unauthorized users gaining access to a computer.

Keywords: biometric authentication, computer security, identity verification, mouse dynamics.

I.INTRODUCTION

User authentication is the process verifying claimed identity. The authentication is accomplished by matching some short-form indicator of identity, such as a shared secret that has been prearranged during enrollment or registration for authorized users. This is done for the purpose of performing trusted communications between parties for computing applications.

Conventionally, user authentication is categorized into three classes:

- Knowledge - based,
- Object or Token - based,
- Biometric - based.

The knowledge-based authentication is based on something one knows and is characterized by secrecy. The examples of knowledge-based authenticators are commonly known passwords and PIN codes. Passwords are the simplest form of user authentication. The object-based authentication relies on something one has and is characterized by possession. Traditional keys to the doors can be assigned to the object based category.

A new category of biometrics that is gaining in popularity is referred to in the literature as behaviometrics (for behavioral biometrics), where analysis focuses on studying the user's behavior while he interacts with a computing system for the purpose of identification [4]–[6]. One interesting example of behaviometrics is mouse dynamics biometrics [5]–[8]. The following Figure 1. Shows the different classification of biometric methods.

Mouse dynamics biometric recognition involves extracting the behavioral features related to the mouse movements and analyzing them to extract a signature, which is unique for every individual, and as such can be used to discriminate different individuals. The main strength of mouse dynamics biometric technology is in its ability to continuously monitor the legitimate and illegitimate users based on their sessional usage of a computer system. This is referred to as continuous authentication. Continuous authentication, or identity confirmation based on mouse dynamics, is very useful for continuous monitoring applications such as intrusion detection.

However, unlike traditional biometric systems, mouse dynamics biometric technology may face some challenges when applied to static authentication, which consists of checking the user's identity at login time. The key challenge is the data capture process, which requires more time to collect sufficient amount of mouse movements for accurate user identity verification [9] than can reasonably be tolerated, or afforded, in a realistic login process.

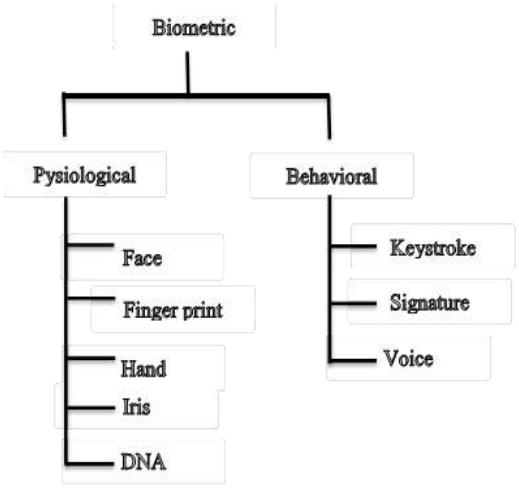


Fig 1. Classification of biometric

We tackle this challenge by proposing we present a new Design and Implementation for authenticating login session Using Mouse Biometrics that allows performing the authentication in a short time and as such may be used for static authentication (at login time). We use mouse gestures to achieve our goal. A mouse gesture results from the combination of computer mouse movements and clicks in a way that the software recognizes as a specific command. In our work, during the enrolment phase the user draws a set of gestures several times on a computer monitor using a mouse along with user details and password. We extract features from the captured gesture, analyse them, store it in database corresponding to particular user profile . In the verification phase, the user will be asked to replicate the gestures drawn by her in the enrolment phase along with password to test against her stored profile .Here we using both password and mouse dynamics for authentication. Even if the imposter get the password of a particular legitimate user he cannot login without knowing gesture .Again gesture depends on someone's behaviour .The system architecture is shown in fig 2. These features are usually unique for a person and cannot be stolen. In Authentication by mouse movements use of any additional hardware isn't required and no one can steal the mouse movements of a person. This is how Biometrics technologies are gaining popularity due to the reason that when used in conjunction with traditional methods of authentication they provide an extra level of security.

The remainder of this paper is organized as follows. In Section II, we summarize and discuss related work. In Section III, we give an overview of mouse biometrics and present the design of our detection system. In Section IV,we present our experimental evaluation process and results. Section V, we give conclusions and summarize our future work.

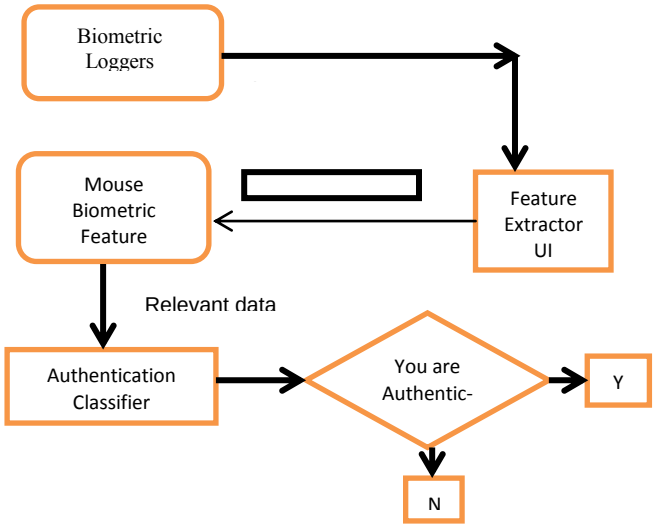


Fig 2 System architecture

II RELATED WORK

Behavioural biometrics first gained popularity with keystrokdynamics with mouse, work on password hardening in 1999 [7]. Later on, Ahmed and Traore [1] proposed an approach combining keystroke dynamics with mouse dynamics. Mouse dynamics for re-authentication have been previously studied as a standalone biometric by Pusara and Brodley [8]. Gamboa et al. [3] performed similar research by conducting an experiment to capture user interaction based on the mouse while playing a memory game. Fifty volunteers participated in the experiment. sequential forward selection technique based on the greedy algorithm was used to select the best single feature and then add one feature at a time to the feature vector. Gamboa et al. [3] showed that the equal error rate (EER) progressively tends to zero as more strokes are recorded. This means that the more interaction data the system records, the more accurate the system should be. But, as we commented earlier, it might be difficult to use such a method for static authentication at login time since the authors reported that the memory game took from 10–15 min to complete on average. so far only three papers in the literature, published by Syukri et al. [6] and Revett et al. [5], Bassam Sayed et.al[9] have targeted the use of mouse dynamics for static authentication. A system that may potentially be used for static authentication, proposed by Syukri et al. [5], utilizes signatures drawn using a mouse for user identification. The extracted features were analysed using geometric average means. The authors conducted two

experiments involving 21 users, in the first of which a static database was used, and in the second a dynamically updated database was used.

Revett et al. [6] proposed a new mouse dynamic analysis approach for static authentication, named mouse lock, which exploits the analogy of a safe, in which the numbers are replaced with graphic thumbnail images. To login, using a mouse, the user is required to click in a password that consists of five images. Bassam Sayed et.al[9] proposed a new mouse dynamics analysis framework that uses mouse gesture dynamics for static authentication. The captured gestures are analysed using a learning vector quantization neural network classifier. Here they combined gestures to get good result.

The main issues with the above works on mouse dynamic are that the minimum amount of data required to achieve meaningful user identification is impossible to obtain within the time constraint of a typical login process. As such, the proposed approaches may be used for user re authentication (after login) or for continuous authentication, but they may not be suitable for static authentication (at login time). And also in Bassam Sayed et.al[9] they have combined gestures to get good result again time consuming .

III. MOUSE GESTURE DETECTION AND ANALYSIS

In this section, we give an overview of mouse gestures and present the design of our gesture detection system.

A. Gesture Creation Module

The gesture creation module is a simple drawing application used to ask the participant to freely draw the gestures. It is used during both enrolment and login phase.

The main purpose of this module is to make the participant draw the gestures in his own way during enrolment and replicate them during login phase. Figure 3.shows example of drawing gesture having 8 data points gestures. It is important to note here that the gestures are not tied to any language and they do not necessarily have a meaning.

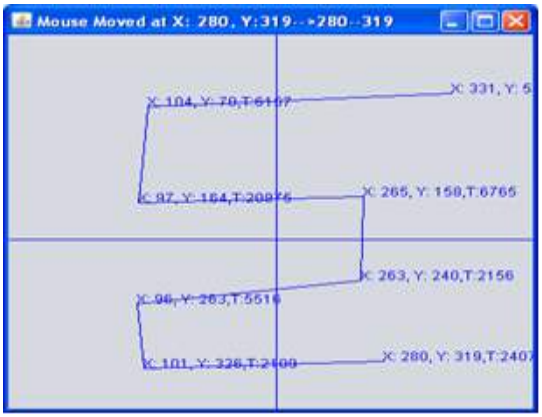


Fig. 3 shows example of drawing gesture having 8 data points gestures .

B. Data Acquisition Module

The data acquisition component loads the gestures, created initially by the user using the gesture creation module. It is used during both enrolment and login phase .The main use is to extract features from the gesture. The data acquisition module records the user interaction while drawing the gesture. The raw data collected from the drawing area consist of the horizontal coordinate (x-axis), vertical coordinate (y-axis), and records the elapsed time t in milliseconds starting from the origin of the gesture. Given gesture can be defined as a sequence of data points and each data point can be represented by a triple <x,y,l > consisting of the X-coordinate, Y-coordinate, and Number of points(clicks) l .The elapsed time in milliseconds starting from the origin of the gesture. During Enrolment phase, for each user, the program creates a directory that will contain the user replications for the different gestures. Each gesture must be replicated a specific number of times (e.g., 10 times). The user has to wait 3 s between consecutive replications. The idea behind this waiting time is to prevent the user from drawing the gesture too fast. The module asks the user to release the mouse between each successful replication during the wait time. We assumed that the wait time and mouse release will force the users to maintain their normal behaviour each time they replicate the gesture.

C. User Registration Module (Enrolment Module)

Registration Module records the user personal information along with gesture, here user supposed to draw gesture n times, the gestures are drawn using gesture creation module. The main purpose of this module is to make the participant draw the gestures in his own way to replicate them later on. It is important to note here that the gestures are not tied to any language and they do not necessarily have a meaning.

They can be any drawing that can be produced in a uni-stroke. Also, this module serves as a practice step for the participants to get familiar with the idea of drawing mouse gestures.

After gestures creation, using the data acquisition the features are extracted from the each gesture replica and stored in Feature data base along with the personal information and user name. All information stored in database are encrypted by using AES Encryption. The below Figure 4 shows extraction of features set from all replicas during enrolment phase.

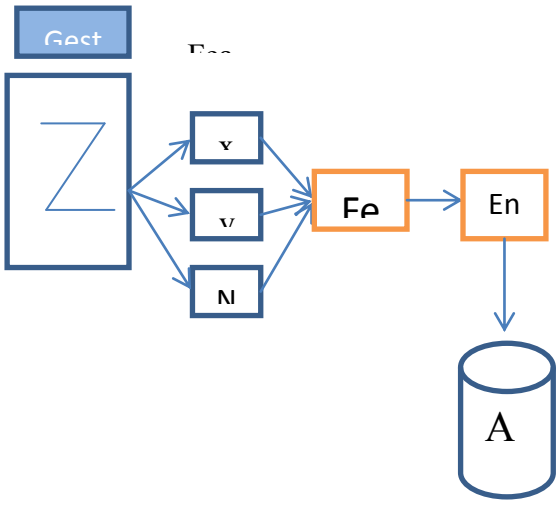


Fig .4 shows the extraction of feature set during login phase.

D. Login Module

In this module, user who is claiming an identity will be asked to replicate gesture he already sketched in the enrolment phase along with user name. Here also by using data acquisition module the features are extracted from the gesture and stored in the authentication data base. The information of the login user is encrypted by using AES Encryption. The idea behind Encryption for login information is just to provide extra security.

E. Authentication Classifier Module

This is the main module which tests the authentication of user. This module decides, given user is legitimate or imposter. For a given user, it exports features gesture replication recorded during the enrolment phase from the feature database and decodes it.

Again it exports the feature of gesture recorded during login phase from authentication database and decodes it. The comparison of gesture in terms of feature takes place .The feature recoded during login phase is compared with each feature recorded during enrolment phase. Before comparison we have to consider few things. Generally humans can not draw the same gesture with the same exact detail twice under the same conditions. This will result in some variability in the replicas produced by the same individual for the same gesture. We have to minimize the effect of variation .This minimization is accomplished by using range normalization. For each point recorded coordinates given the relaxation of +20 and -20 units .The idea behind this is user can't draw the gesture exactly with same position for all points, instead we given the space for each point .user can draw the gesture within that space for ach point.

AS specified each feature recorded during login phase is compared with features of all replicas during enrolment phase. First it compares with number of points (clicks) ,then it compares with x,y coordinates considering range relaxation. While comparing coordinates, it takes first point coordinate form login gesture and compares with same point coordinates of all gesture replicas. At least, in any one of the gesture in replicas the coordinate should match. The comparison is repeated for all the points in gesture .The decision depend on the total matching count of all the points in gesture. The matching count should be greater than or equal to Number of points in the gesture .

IV EXPERIMENTAL EVALUATION

We present, in this section, the experimental evaluation of the proposed framework. We start by describing the experimental conditions and procedures, and then present, analyse, and discuss the obtained results.

The main objective of our experiment was to be able to recognize individuals based on their mouse gestures. Ideally, the system should be able to recognize, with a high degree of accuracy, the behaviour of each user while replicating a specific gesture were involved in our experiment. So in our approach, we used the gestures having the combinations of angles, lines, curves. The main reasoning for this choice was, the more angles and curves the gesture has, the more it will require muscle tension and concentration from the users. This, in turn, imposes the intrinsic behaviour of the human motor control while drawing such gestures.
To show the efficiency of the proposed system ,two users user1 and user2 are asked to draw the same gesture the two replicas produced by user1 is

compared with replica produced by user2 as shown in below Figure 7.

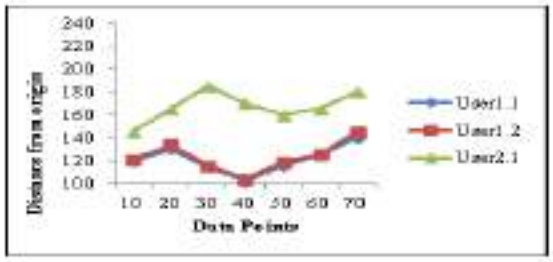


Fig..5 Comparing distance from original features of two replicas belonging to User 1 and one replica belonging to User 2 for the same gesture.

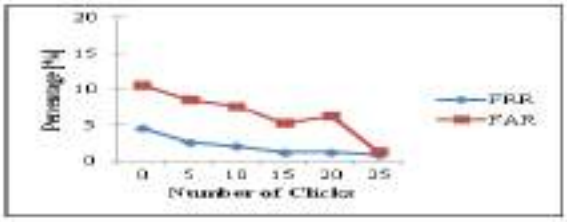


Fig .6 FAR and FRR of a Gesture

Even though both the user drawn the same gesture but replica produced by both user are different the more angles and curves the gesture has, the more it will require muscle tension and concentration from the users also as show in figure 8 False Acceptance Rate(FAR) and False Rejection Rate decreases for gesture having More Clicks. This, in turn, imposes the intrinsic behaviour of the human motor control while drawing such gesture.

V CONCLUSION AND FUTURE STUDY

In this paper, new Design and Implementation for authenticating login session Using Mouse Biometrics. Mouse dynamics are defined by the characteristics which are acquired by analysing the inputs a user performs by a pointing device. These characteristics are then combined to form the factors which are stored in the user profile. During Login session, comparing the current input stream with the previously stored user profile, mouse dynamics can detect anomalies in the user's behaviour and stop intrusions while they are happening. One of challenge faced by most of security system are protection against security attacks. Like many other biometric technologies, mouse dynamics can be the target of replay attacks. In our proposed system Such threats are mitigated by strengthening the protection of the biometric information using Encryption techniques.In future work, we intend to enhance the accuracy of our proposed scheme by revisiting various aspects. Since our proposed system is entirely software based, integration in a complex system environment .During comparison we give range Normalization (-20 and +20) units which may reduce by Using advanced software would allow addressing the interoperability challenges inherent in complex system environments. Mouse dynamics can also be the target of automated attacks, also referred to as generative attacks, where high quality forgeries can be generated automatically using a small set of genuine samples. We plan, in our future work, to strengthen our system by investigating the impact of generative attacks against it.

REFERENCES

[1] S. Cimato, M. Gamassi, V. Piuri, R. Sassi, and F. Scotti, "Privacy-aware biometrics: design and implementation of a multimodal verification system," in Proc. Annu. Comp. Sec. Apps. Conf., 2008, pp. 130–138.

[2] D. Lopresti, F. Monrose., and L. Ballard, "Biometric authentication revisited: Understanding the impact of wolves in sheep's clothing," in Proc. 15th USENIX Sec. Symp., 2006.

[3] H. Gamboa and A. Fred, "A behavioral biometric system based on human-comp. inter," in Proc. Conf. Biometric Tech. Human Identification, vol. 5404. 2004, pp. 381–392.

[5] K. Revett, H. Jahankhani, S. de Magalhaes, and H. M. D. Santos,"A survey of user authentication based on mouse dynamics," in Proc.ICGeS, CCIS'12, 2008, pp. 210–219.

[6] A. F. Syukri, E. Okamoto, and M. Mambo, "A user identification system using signature written with mouse," in Proc. 3rd Australasian Conf. Inform. Sec. Privacy, 1998, pp. 403–414.

[7] C. Varenhorst. (2004). Passdoodles: A Lightweight Authentication Method [Online]. Available: http://people.csail.mit.edu/emax/ papers/varenhorst.pdf

[8] M. Gamassi, M. Lazzaroni, M. Misino, V. Piuri, D. Sana, and F. Scotti, "Accuracy and performance of biometric systems," in Proc. Instrum. Meas. Tech. Conf., 2004, pp. 510–515.

[9] Bassam Sayed Bassam Sayed, Issa Traor´e, Isaac Woungang, and Mohammad S. Obaidat, *Fellow, IEEE* l Biometric Authentication Using Mouse Gesture Dynamics.

Secured Secret Image Transmission by using Fragment Visble Mosaic Image Technique

Ashalatha H M, M.D Anitha Devi,Dr. K.B Shivakumar, Dr. M. Z Kurian
Dept. of ECE,Sri Siddhartha Institute of Technology ,Tumkuru Karnataka, India
Ashalatha401@gmail.com ,anumdssit@gmail.com, shivakumarkb@ssit,edu.in, mzkurianvc@yahoo.com

ABSTRACT : In recent years as many of our daily tasks are related to internet, the security of information being transmitted over the internet is an important issue. Many applications or fields require transmitting secret image over the internet that may contain private and confidential information that is needed to be protected from hacker or from unintended users. Many techniques and methods have been proposed to ensure security, out of them most commonly used techniques are encryption and data hiding. Encryption ensures security but the problem with encryption algorithm is that the resultant noise image attracts the attention of hacker so it may be possible after many trials they can be decrypted by hacker. Another method is data hiding in which pieces of secret information is hidden behind a carrier that may be anything a text file, video, audio or an image. Now a day's, for secure image transmission a new concept that is of mosaic image is used in the field of data hiding. A new type of mosaic image that is called as secret fragment visible mosaic is formed to hide the existence of secret image by using a carrier image. In this project a new technique for creating a secret fragment visible mosaic image is proposed.

Keywords – Mosaic image, secret fragment visible mosaic image, reversible color transformation, data hiding and steganography algorithm.

I. INTRODUCTION

In recent years the secret information is hided in the images from various sources and are transmitted through the internet in many applications. Medical and military imaging systems and databases are examples of such applications. These images consist of some private or confidential data which is to be Kept confidential and protected from leakage process while transmitting. Many techniques are available to provide security for image transmission operations. There are two general methods they are image encryption approach and data hiding technique.

The image encryption uses images which have a neutral property like high redundancy and strong correlation. There is a possibility that the image which is created using image encryption technique may be attracted by the third party because of its randomness in nature while transmission. By using these properties of the image researchers worked on authenticating the image. Image Encryption algorithms make use of these natural properties of the image to authenticate the image. Another method of authenticating the color image is data hiding where we use two types of entities to transmit the image secretly. One is an image which we required to transmit secretly called as secret image and another one is an image which is used to hide the secret image called as a cover image.

These encryption and decryption process is controlled by key at transmitter and receiver end. Without the key we cannot decrypt the image at receiver end. Several data hiding techniques have been proposed in the literature includes LSB Substitution [1], histogram Shifting [2], difference expansion [3], prediction error expansion [4]. In this paper we have implemented an authentication system for an image which transforms a secret image into meaningful mosaic image which looks like a preselected carrier image. We are using a simple LSB Substitution to hide the data required for recovering secret image at receiver end. The method implemented in this paper is inspired by Lai and Tsai [5] and Lee and Tsai [6]. The mosaic image is the result of rearrangement of the fragments of a secret image in disguise of another image called the carrier image preselected from a preselected database. Specifically, a secret image and a carrier image first divided into rectangular fragments and then the secret image blocks is fit into carrier image P a g e blocks according to a similarity criterion based on color variations. Next, the color characteristic of each secret image block is transformed to be that of the corresponding carrier block in the carrier image, resulting in a mosaic image which looks like the carrier image. The relevant information required for recovering the original image id hide into the created mosaic image. The image encryption algorithms yield a mosaic image which is meaningless. The data hiding method must be hide data in highly compressed manner into a disguising mosaic image without compression.

II. RELATED WORK

The authorsYa-Lin Lee ; Nat. Chiao Tung Univ., Hsinchu, Taiwan; Wen-Hsiang Tsai, proposed a paper titled as "A New Secure Image Transmission Technique via Secret-Fragment-Visible Mosaic Images by Nearly Reversible Color Transformations". It is a new secure image transmission technique, in which the given selected large volume secret image is transformed into a mosaic image called as secret-fragment-visible mosaic image of the same size as that of secret and target image. The created mosaic image looks similar to a selected target image and it may be used as a disguise of the secret image, which is constructed by dividing the secret image into small fragments and by color transforming those fragments color characteristics to be those of the corresponding blocks of the selected target image. There are various color transforming techniques which are designed skillfully to conduct the color transformation process by doing this the secret image may be recovered lossless. In this paper the author proposed a new scheme of handling the overflows/underflows. The information which is needed to recover the secret image is embedded into the created mosaic image by a lossless data hiding technique with a secret key.

The author Merlin et all. Proposed a new method for secured communication in the paper titled as "Covert Image Transmission Technique Using Mosaic Image". This is a novel method for secret communication which involves transmission of secret images. The color image which is to be transmitted secretly is disguised into a cover image of the same size which produces a mosaic image. The mosaic image looks as mirror image of the cover image and is visually indistinguishable from it. The creation of mosaic image involves block by block processing of both secret and cover images. In the proposedmethod the Gaussian noise is added to the secret Image to ensure positive variance of intensities within image blocks. Image blocks are matched according to the standard deviation of the intensities. Color transformation technique is applied to transform the secret image blocks into theblocks of mosaic image. The information required which is required to retrieve the secret image from mosaic image is compressed and embedded in the mosaic image by using a LSB embedding technique. The RMSE and PSNR ratios are considered to analyze the performance of the method.

I-Jen Lai and Wen-Hsiang Tsai, Senior Member, IEEE proposed a novel method for secure communication in the paper titled as "Secret-Fragment-Visible Mosaic Image–A New Computer Art and Its Application to Information Hiding". It is a new type of computer art image which is called as secret-fragment-visible mosaic image, so called mosaic image can be created by dividing the selected secret image into a small fragments and embedding them in to a cover image to form a mosaic image of a given image to become a target image in a mosaic form, this achieves an effect of embedding the secret image visibly but secretly in the resulting mosaic image. This method of hiding a secret image is useful for covert communication or secure keeping of secret images. The author proposed a method for transforming the 3-D color space to 1-D color scale to transfer the secret image and to create a mosaic image and a new image similarity measure is proposed for the selection of target image from the database, based on which a new image similarity measure is proposed for selecting from a database. A fast greedy search algorithm is proposed to select a secret image tile to fit into corresponding block in target image. The information regarding the tile image fitting sequence is embedded into randomly-selected pixels in the mosaic image using a lossless LSB replacement technique with the secret key. Good experimental results show the feasibility of the proposed method.

III. PROPOSED WORK

The block diagram of the proposed method is as shown in figure 1 In Proposed System Secrete image is pre-processed before caring out mosaic image process, in pre-processing both secretes and target images are resized to a standard image resolution, It will avoid the dimensionality error during the processes. The Images are divided into blocks or tiles; from each block standard deviation is calculated and stored it in a ascending order. Using sorted information first tile of secrete mage is fit to particular first block of Target similarly all the image tiles are fit in Target image and resulting image looks like Selected Target. Then Color Transformation is done to Mosaic Image, image tiles are rotated using RMSE value to get final Mosaic Image. In the above block diagram fig.1 Shows the creating of mosaic image using secrete and target image.

Steps involved in generating secret fragment visible Mosaic image are given in this section:

A. Selection of cover and payload images:

In this step user has to select their secret image and any random image of his/her choice as cover image user can select any image as target image of any size in contrast to Lai and Tsai method that requires cover image should be double in size and should satisfy some similarity, measure criteria to be used s cover image. But in order to avoid to suspicion it is advised that target image is to be selected, which is of some field or of same background as that of secret image.

B. Resize the secret image:

Here in this proposed method to create secret fragment visible Mosaic of the same size as that of cover image, after the payload image and cover image are selected, we have to check whether the cover image and payload image are of same size or not, if they are not of same size then we have to resize the payload image to make it of same size as that of cover image.

C. Divide both the payload and cover images into tiles:

Next step is to split the source image into small pieces called as tiles. In proposed algorithm in order to create the secret fragment visible mosaic image there is requirement that the number of blocks of target image should be same in size. So the cover and payload images are to be split by using splitting technique.

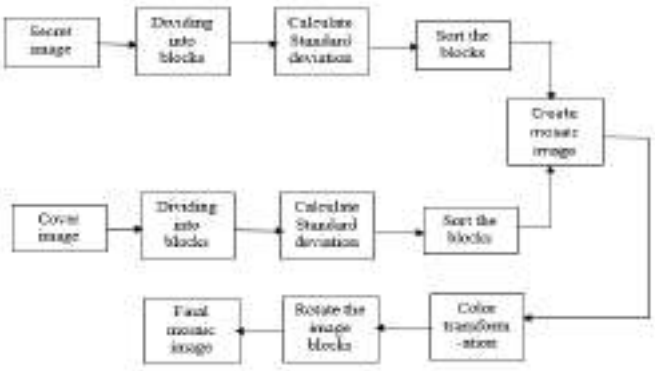


Fig 1: proposed system.

D. Calculate standard deviation and arrange them in ascending order:

Here how to choose an appropriate block B of target image for each of the tile T of secret image is the main issue. For this the standard deviation of the pixels of each block is to be calculated as a similarly measure value to select of the tile T of payload image .Then the tile image is found by arranging the tiles in ascending order to form the sequence stile and all the blocks to form sequence target that is in

ascending order with respect to their standard deviation value.

E. Creation of Mosaic image:

To create the Mosaic image, the used sorted sequence of standard deviation to form resultant image. First we fit the first tile in sequence stile into the first block in sequence target, and accordingly fit the second tile in stile to second block in target and so on. In this way resultant Mosaic image is formed which will look somewhat similar to the selected cover image.

F. Perform Color Transformation:

The Mosaic image created above contains some noise which willdegrade the picture quality of the Mosaic image so to make the Mosaic image to look very similar to the cover image color transformation has to be applied to the Mosaic image. Calculate the RMSE value for each block by rotating the each blocks to the angles of 0^o, 90^0, 180^0, 270^0 and the blocks are rotated to the angle with the least RMSE value. The Mosaic image created after the color transformation will look same as that of cover image.

G. Reconstruction of Secret image with the Mosaic image :

Reconstruct the secret image with the Mosaic image by carrying de-embedding process.

IV. RESULTS AND ANALYSIS

Secret Fragment Visible Mosaic image creation involves following steps:

A. Selection of cover and payload images

B. Resize the secret and cover images

C. Divide both the payload and cover images into tiles

D. Calculate standard deviation and arrange them in ascending order

E. Creation of Mosaic image

F. Perform Color Transformation.

The Mosaic image created with the above steps is as shown in fig. 2 and fig. 3 it looks very similar to the selected target image.Inputs to the Mosaic image creation are target image and secret image. Simulation results of mosaic image creation are shown below. Mosaic image which is created with the target and secret images and the reconstrucred secret image from Mosaic image are shown in fig. 2 Secret iamge is retreived back without any loss of information.

Fig 2: simulation results of Mosaic image constructed with selected target andsecret image and secret image retrieved from the Mosaic image.

Fig 3: simulation results of Mosaic image constructed with different target and same secret image and secret image retrieved from the Mosaic image.

V. CONCLUSION

In this paper a new method for creation of secret fragment visible Mosaic image has been proposed. This method allows user to select any cover image of their choice to create Mosaic image and also user can select payload image and cover image of any size to create Mosaic image. The resultant Mosaic image can be used for secret communication and the secret image is retrieved from the Mosaic image (Fig. 2 and 3) without any loss. This method of creation of mosaic image has the advantage of strong secret data embedding capacity so the unintended user cannot decrypt the secret data.

REFERENCES:

[1] C. K. Chan and L. M. Cheng, Hiding data in images by simple LSB substitution,‖ Pattern Recognition..,vol. 37, pp. 469474, Mar. 2004.

[2] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, Reversible data hiding,‖ IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 3, pp. 354362, Mar. 2006.

[3] J. Tian, ―Reversible data embedding using a difference expansion,‖ IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 8, pp. 890896, Aug. 2003.

[4] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y.-Q. Shi, Reversible water- marking algorithm using sorting and

prediction,‖ IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 7, pp. 989-999, Jul. 2009.
[5] I. J. Lai and W. H. Tsai, Secret-fragment-visible mosaic image—A new computer art and its application to information hiding,‖ IEEE Trans. Inf. Forens. Secure.,vol. 6, no. 3, pp. 936–945, Sep. 2011.
[6] Ya-Lin Lee and Wen-Hsiang Tsai, A New Secure Image Transmission Technique via Secret-Fragment Visible Mosaic Images by Nearly Reversible Color Transformations,‖ IEEE Trans. Circuits and systems for video technology., volume:24 , Issue: 4 ,pp. 695 – 703, April 2014.

A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction

Deepu S[1], Pethuru Raj[2] and S.Rajaraajeswari[3]
[1]2nd Year MCA, Raja Rajeswari College of Engineering, Bangalore-560074, India
[2]Infrastructure Architect, IBM Global Cloud Center of Excellence, IBM India, Bangalore-560045, India
[3]Department of MCA, Raja Rajeswari College of Engineering, Bangalore-560074, India

ABSTRACT- With the steady accumulation of unstructured data, the domain of natural language processing (NLP) is gaining widespread attraction amongst researchers and practitioners in order to quickly and easily extracts prediction-like insights in a simplified and streamlined fashion. The subject of text mining and analytics is going through a variety of delectable advancements. There are a number of articulations on the subjects of NLP and machine learning (ML). Very recently, the model of the bag of words has become so popular in order to produce accurate predictions out of unstructured text data. In this paper, we have explained an easy-to-use framework for accelerated usage of the BoW model towards pioneering text mining and processing. We have demonstrated a simple example by leveraging this framework in order to showcase the utility of this generic framework that can be easily replicated across in many other associated scenarios.

Index Terms—Bag of Words, Machine Learning; Natural Language Processing, Text Mining, Predictive Analytics, Bernoulli Document Model.

I. INTRODUCTION

Predictive analytics has become the key subject of study and research these days [1]. As data become big data due to the explosion of unstructured text documents, the researchers across the globe are focusing on unearthing and experimenting a variety of powerful algorithms and approaches for speeding up the process of processing online as well as off-line text files with the aim of producing something useful to act upon with all the clarity and confidence. There are algorithms in plenty for doing various analytical activities in order to extract actionable insights in time out of big, fast, streaming and IoT. Machine learning [2] is emerging as one of the hot topics in the industry as well as in academic institutions to smoothen the analytical tasks.

With the steady growth of text content, the paradigm of natural language processing (NLP) [3] is fast-evolving these days. Inspired minds and luminaries are striving hard and stretching further to bring forth a bevy of workable methods and mechanisms that in turn facilitate pragmatic knowledge and predictions. We have found through our extensive experiences that the emerging BoW model seems to be a better fit for predicting something by processing text content. This paper is for explaining the nitty-gritty of the proposed framework for expediting the prediction from large-scale unstructured text data.

II. EXPLAINING THE BAG OF WORDS PREDICTION MODEL

The Bag of Words (BoW) model learns a vocabulary from all of the documents, and then models each document by counting the number of times each word appears. The BoW model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity. Recently, there are different domains showing a lot of interesting in leveraging this BoW technique towards efficient text analytics. The BoW model is commonly used in methods of document classification, where the (frequency of) occurrence of each word is used as a feature for training a classifier [4, 5, 6].

For example, consider the following two sentences:

Sentence 1: "The cat sat on the hat"
Sentence 2: "The dog ate the cat and the hat"

From these two sentences, our vocabulary is as follows:

{the, cat, sat, on, hat, dog, ate, and }

To get our bags of words, we count the number of times each word occurs in each sentence. In Sentence 1, "the" appears twice, and "cat," "sat", "on", and "hat" each appears once, so the feature vector for Sentence 1 is:

{the, cat, sat, on, hat, dog, ate, and }

Sentence 1: {2, 1, 1, 1, 1, 0, 0, 0}

Similarly, the features for Sentence 2 are: {3, 1, 0, 0, 1, 1, 1, 1}

For example, as per the Wikipedia, the BoW model has also been used for computer vision. In computer vision, the BoW model can be applied to image classification, by treating image features as words. In document classification, a bag of words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.

III. THE BOW BASED ANALYTICS METHODOLOGY

Text classification is the task of classifying documents by their content: that is, by the words of which they are comprised. Perhaps the best-known current text classification problem is email spam filtering: classifying email messages into spam and non-spam. We have arrived at and articulated an optimal methodology for executing the BoW model on text data sets based on our wide experiences gained out of various works that intrinsically leveraged the distinct power of BoW technique. The steps are enumerated below.

1. **Reading the Unstructured Text Data –** The unstructured text data that needs to be subjected to a variety of investigations. So the text files need to be download and ingested. For example, there are tweets, Facebook comments, blogs, the addresses of celebrities in various fields, the various sentiments expressed by users, etc. We have used R Studio to read the file to be examined.
2. **Data Preprocessing** – This is a phase wherein all kinds of noises of text data need to be filtered out. If there is any kind of markers and extra white spaces, they need to be identified and eliminated before jumping into the processing and analytics phase. If there is any HTML tags or punctuation marks, numbers, etc. in the text document, they need to be meticulously identified and erased in order to arrive at highly organized and optimized document. There are other cleaning operations to be accomplished through a number of reviews and refinements. There are already developed libraries, packages, and classes in order to remove all kinds of
3. **Knowledge Discovery** – As articulated in the beginning, the Bag of Words model learns a vocabulary from all of the documents and then models each document by counting the number of times each word appears. We have explained this phase in detail through a practical example below in order to enhance the readability and understandability of the BoW technique in extracting predictions out of massive scales of text data. We have chosen to use the Bernoulli Document model [7], which is detailed below.
4. **Knowledge Dissemination** – There are visualization platforms, dashboards, report-generation tools, etc. for displaying and demonstrating the extracted in a user-preferred manner. There are automated systems to showcase the knowledge in a 360-degree view.

This is a very generic framework and there are multiple technologies, tools, and techniques to be leveraged. Advanced algorithms can be easily sneaked in to ensure high performance and other non-functional requirements.

IV. ABOUT THE BERNOULLI DOCUMENT MODEL

Text classifiers typically don't use any kind of deep representation about the language. A document crafted in the language is represented as a bag of words. A bag is like a set that allows repeating elements. This is an extremely simple representation. That is, it only knows which words are included in the document and how many times each word occurs. This does not take the word order into account. There are two probabilistic models of documents (multinomial model and multivariate Bernoulli model). Both represent documents as a bag of words using the Naive Bayes (NB) assumption. Both models represent documents using feature vectors whose components correspond to word types. If we have a vocabulary V, containing |V| word types, then the feature vector dimension is d=|V|.

A. Bernoulli Document Model

A document is represented by a feature vector with binary elements taking value 1 if the corresponding word is present in the document and 0 if the word is not present. Bernoulli model uses binary occurrence information, ignoring the number of occurrences, whereas the multinomial model keeps track of multiple occurrences.

B. A Sample Application Development Steps

1) Installing R and RStudio

RStudio is an integrated development environment (IDE) for R. We have used the free version. Like R, RStudio installs painlessly and also detects your R installation. Text mining and certain plotting packages are not installed by default so one has to install them manually. The relevant packages are:

1. **tm** – the text mining package.
2. **SnowballC** – required for stemming
3. **ggplot2** – plotting capabilities
4. **wordcloud** – which is self-explanatory

The simplest way to install packages is to use RStudio's built-in capabilities (go to *Tools* > *Install Packages* in the menu).

2) Dataset Ingestion

For our experimentation, we have taken a data set from this link [9]

https://archive.ics.uci.edu/ml/datasets/Bag+of+Words

Download and ingest the dataset into R Studio. There is a provision for file ingestion.

3) Cleansing for Test Processing, Mining and Analysis

The text file has to go through a series of corrections such as the removal of numbers, capitalization, common words, punctuation and other noises as the first and foremost step for subjecting text documents to a variety of investigations.

4) Leveraging Bernoulli Document Model

By using the text mining software, which has intrinsically implemented the Bernoulli document model, it is possible to bring forth word matrix.

5) Plotting the Obtained

There are additional software modules such as ggplot2 and wordcloud to create effective visualizations.

V. THE PICTORIAL REPRESENTATION OF METHODOLOGY

The prominent steps are being illustrated through the following diagram.

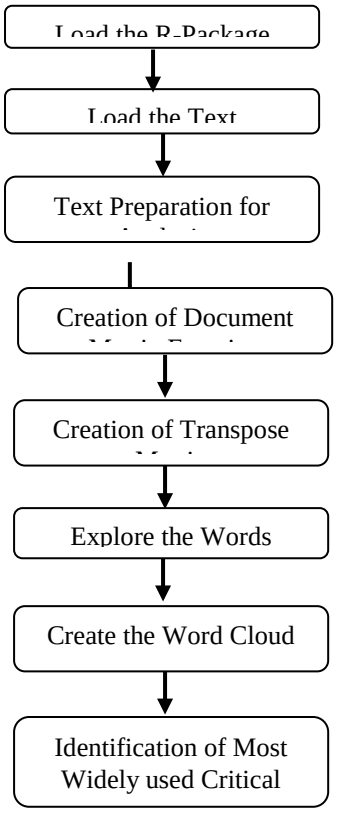


Figure 1 – The Process Flow Chart for Text Mining using "R"

VI. THE SAMPLE OUTPUT

With the use of wordcloud solution, it is possible to craft different outputs as evidently indicated below.

Figure 2 – The Word cloud Output

The ggplot2 is capable of producing the following diagram indicating the words that appear more than 5 times

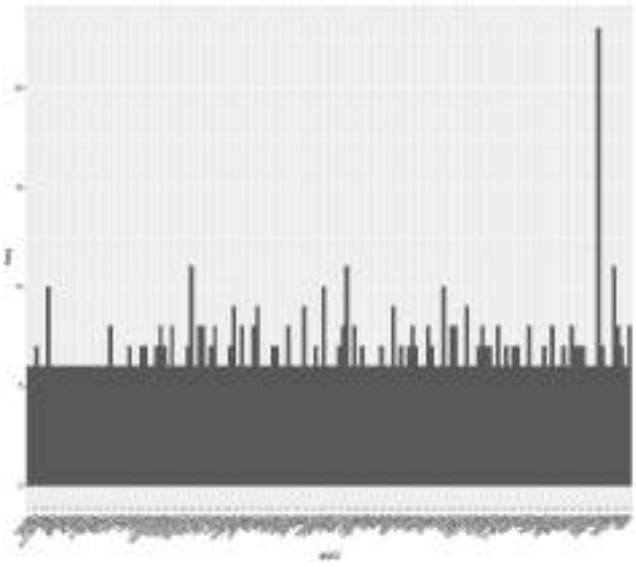
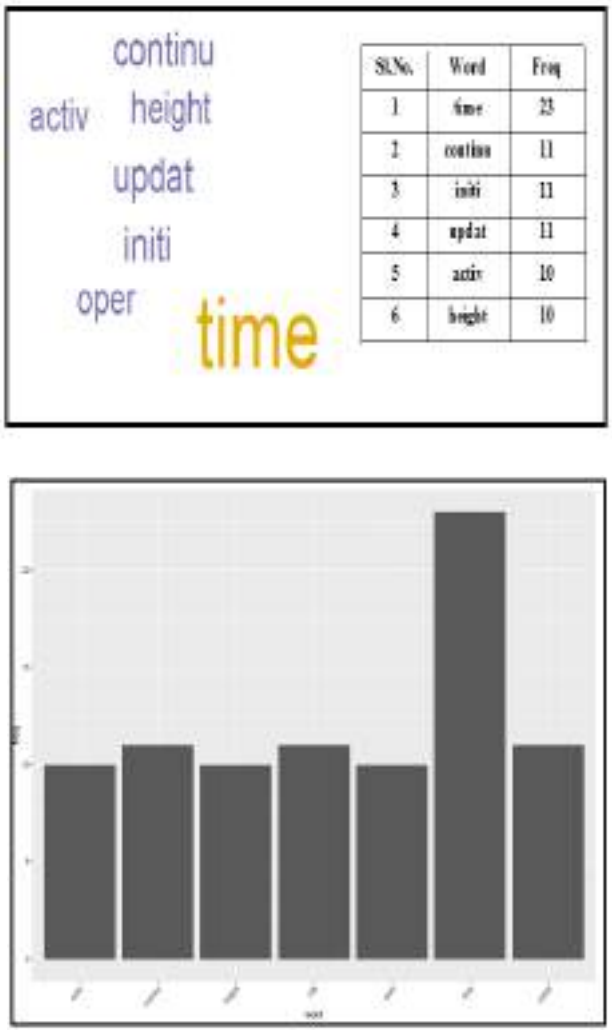


Figure 3 - Words Appeared More Than Five Times

The next output indicates the number of words that appear

more than 10 or more times.



Sl.No.	Word	Freq
1	time	23
2	continu	11
3	initi	11
4	updat	11
5	activ	10
6	height	10

Figure 4 – The Words Appeared More Than Ten Times

VII. CONCLUSION

We all know that data heaps are blessed with a variety of pragmatic knowledge. However, there are two major barriers. There is a rapid growth of unstructured text files originating from different sources and social sites and the amount of data getting generated, captured, and subjected to purpose-specific processing, mining and analytics are growing exponentially. In order to extract any hidden patterns among text files, there comes a number of ground-breaking algorithms, game-changing software solutions, integrated platforms, NoSQL databases, and cloud infrastructures. In this paper, we have brought in an enabling framework to make sense out of word documents quickly and easily. We have also demonstrated the efficiency and effectiveness of our framework through a sample implementation. There are a few BoW models and we plan to do a comparative study of them before incorporating the best-in-class BoW model in order to strengthen our framework.

REFERENCES

[1] Frank Buytendijk and Lucie Trepanier, (2010), Predictive Analytics: Bringing the Tools to the Data, Oracle Whitepaper
[2] Pedro Domingos, A Few Useful Things to Know about Machine Learning, *https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf*
[3] Ronan Collobert, JasonWeston, L´eon Bottou, Michael Karlen, Koray Kavukcuoglu and Pavel Kuksa, (2011), Natural Language Processing (Almost) from Scratch, Journal of Machine Learning Research, 12, 2493-2537
[4] Dani Yogatama and Noah A. Smith, (2014), Making the Most of Bag of Words: Sentence Regularization with Alternating Direction Method of Multipliers, Proceedings of the 31st International Conference on Machine Learning, Beijing, China, JMLR: W&CP, volume 32
[5] Cordelia Schmid, Bag-of-features for category classification, *www.cs.umd.edu/~djacobs/CMSC426/BagofWords.pdf*
[6] Jialu Liu, Image Retrieval based on Bag-of-Words model, *jialu.cs.illinois.edu/technical_notes/CBIR_BoW.pdf*
[7] Hiroshi Shimodaira, (2015), Text Classification using Naive Bayes, Learning and Data Note 7, Informatics 2B
[8] https://cran.r-project.org/mirrors.html
[9] https://archive.ics.uci.edu/ml/datasets/Bag+of+Words

A New Approach of Image Steganography Technique for Information Hiding using Nearest Filling Technique

S. Sreedhar Kumar, Associate Professor, Bhagyalakshmi, (M.tech, CNE)
Department of Computer Science, DBIT, Bangalore.
sree_me_261177@yahoo.co.in, bhagyachandrahasa053@gmail.com,

ABSTRACT- Steganography is an art that involves secret communication by using encryption and decryption from sender to receiver through message, images, videos etc. In this work, A new Approach of image Steganography technique is proposed, This proposed Work aims to, The secret information can be concealed in 0 to N Blocks of image. Hence the user can be achieved high Secure communication. This proposed work consists of two Stages namely Encryption stage, decryption stage. The Encryption stage consists of four steps, in the first step the text message has to convert into ASCII code. In the Second step splitting the image into number of blocks using non overlapping method. In the In Third step, We Introducing a new Method called Nearest Neighbor Filling Method (NNM) aims to hide the converted secret message into the nearest value of image block in linear manner. In the fourth step send the encrypted message with a secret key to the receiver. The receiver stage consists of four steps, in the first step the secret message has to decrypt with a secret key, In the second step the receiver has to find out the input information in the image, the third step is extracts the secret message from the image block, the final step is converting the ASCII code into the original message. Experimental result shows that the proposed approach is Simple and Better suitable to send the secret message in the stego- image.

Key words: concealed, linear manner, nearest neighbor filling, non-overlapping, overlapping.

I. INTRODUCTION

The need of security is very essential in digital transmission of information or data through internet. Steganography is applied computationally, when wrapping of works such as text files, image files, audio files and video files are used here that a secret message can be hidden inside them.[1-3]. Images are preferred medium for the current Steganography techniques. content adaptability, resilience, there exists a large number of image Steganography techniques which are accompanied by various attacks
on the Steganography systems. Security of any Steganography technique depends on the selection of
pixels for embedding.[4], There are two approaches are used to protect secret information from intruder or being attacked by others during transmission. One is encryption which is in the form of encoded in another from by using a secret key before sending the information, which we can called as cipher text. This information can be decoded by using secret key. There are several popular encryption techniques namely, AES, DES, RSA, Blowfish, Two fish etc.
there is an another way is Steganography which is Steganography is an art that involves secret communication by using encryption and decryption from sender to receiver through message, images, videos etc. Steganography technique can be used in military-defense, commercial, on-line activities, anti-criminal so on. There are many Steganography techniques available in the current technological field.[1-10] A new approach of image Steganography technique for information hiding using nearest neighbor filling method(NNM). This approach will be simple and gives a better Result than previous techniques since all works are done by using overlapping method but here introducing a new called finding the corresponding ASCII value for information and hide it into the image blocks by using nearest neighbor filling method(NNM) by linear passion. Hence it will give a better result than previous because there should not be confliction to the user of overlapping of image blocks.

II. PROPOSED SYSTEM

In this section details of the proposed system presented the proposed system consists of 2 stages Encryption and decryption fig.1 shows that the steps involved in proposed system. As per system architecture the sender having plain text that consists of alphabetical sentences or the numeric values or any special characters Ex: welcome to the Steganography world or 12345 or $%#$%^& ,these can be converted into respective ASCII value then ,at the same time, sender needs to have a image that is used to hide these ascii values for the secret communication the image has to split into blocks using non-overlapping method by linear manner that each block is having its own pixel values that involves in the Steganography then, by using vector calculation method and minimum distance formula as shown below, Here the block zero will be assumed that contains a value like

70	100	120	140
90	76	40	30
45	98	34	33
34	55	54	88

Here apply a minimum value calculation as the algorithm used.

$Min\{d(P_i,B_{ij})|\square B_{ij}\epsilon B_i,\square B_i\epsilon B,P_i\epsilon P|\}$ ……..1

Where P_i represents the i^{th} character ascii value B_{ij} is the i^{th} block that belongs to the Image and B_{ij} represents the j^{th} pixel value in the i^{th} block that is in image $d(P_i,B_{ij})$represents the distance between the i^{th} character ascii value with j^{th} pixel in i^{th} block and in distance as minimum value.

$D(P_i,B_{ij})=\{|P_i-B_{ij}|p_i\epsilon P,B_{ij}\epsilon B_i,B_i\epsilon B\}$…….2

- Here i=0,j=0,k=0

D(65,70)=Min|df(65-70)|
Min value=5.

- $D(P_i,B_i)$=Min $|df(p_i-B_{ijk})|$

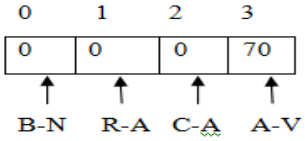
Here i=0,j=0,k=0
D(100,70)=Min|df(100-70)|
Min value=30

- $D(P_i,B_i)$=Min $|df(p_i-B_{ijk})|$

Here i=0,j=0,k=0
D(120,70)=Min|df(120-70)|
Min value=50

- $D(P_i,B_i)$=Min $|df(p_i-B_{ijk})|$

Here i=0,j=0,k=0
D(140,70)=Min|df(140-70)|
Min value=70
Here the value is



B-N:-Block Information
R-A:-Row Address
C-A:-Column Address
A-V:-Actual value

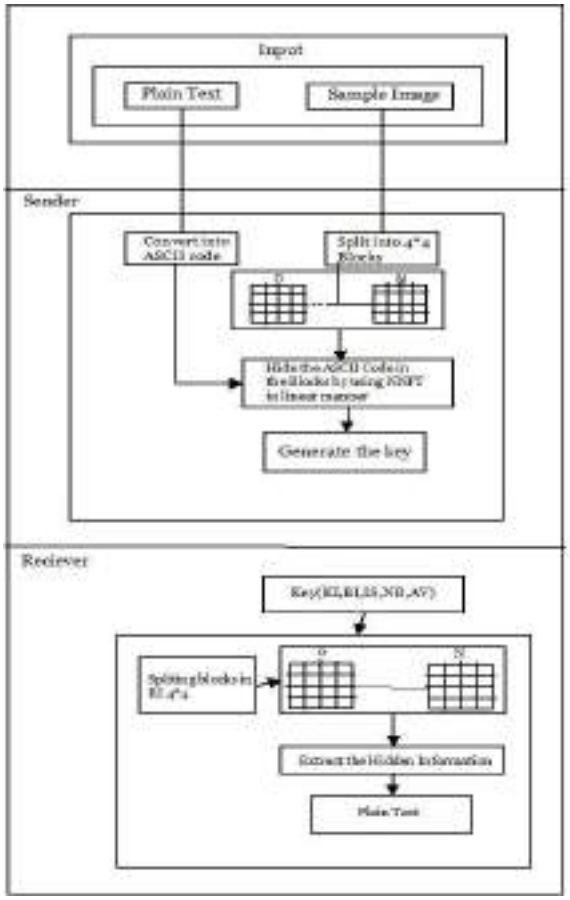


Fig. 1 System Architecture

So here got a nearest value as 70 is the nearest to 65 hence Replace 65 with 70 so encryption has done. Sender send this coded information to receiver with these key values. Receiver receives the message with key. Split the K-I into values then Extract the hidden value Get the ASCII value of each character Get a original information or message.

2.2 Algorithm:

2.2.1 Encryption Process
Input: Plain Text, Image
Output: Ascii Code , Encrypted image
Begin
Sender:
Step 1: Plain text converted to ascii code
Step 2: Fetch the original image.
Step 3: Split the image in to 0-N Blocks by using non-overlapping method.
Step 4: Place the ascii value in to the image blocks by using nearest filling method.
Step 5: Send the encrypted message with key to the receiver side.
2.2.2 Decryption process
Receiver:
Step 1: The receiver decrypts the message with the key.
Step 2: Find the input information in the image.
Step 3: Then extract the secret message from the image blocks.
Step 4: Convert the ascii code into original message.
End

III. EXPERIMENTAL RESULT

The experimental result climes the analysis that the implementation can be done by using java, here is the text message result shown below.

Example

- Input

Sample message-> ABC

- Output

ABC=656667
WELCOME=119101108111109101
so that result is as shown below
The Fig 2 shows the original image before get into the encryption,

Fig 2 shows the original image before encryption

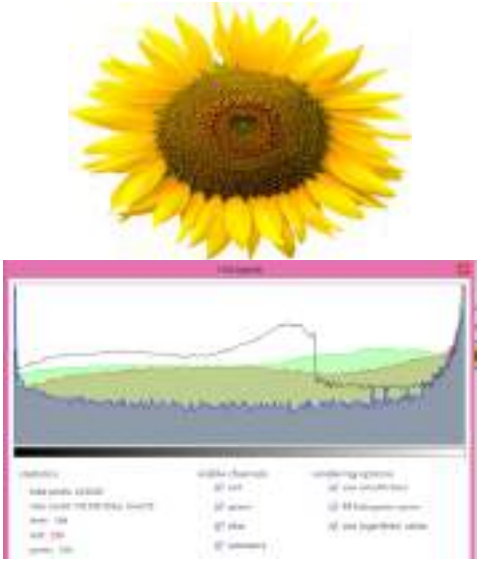


Fig 3 shows the encrypted image

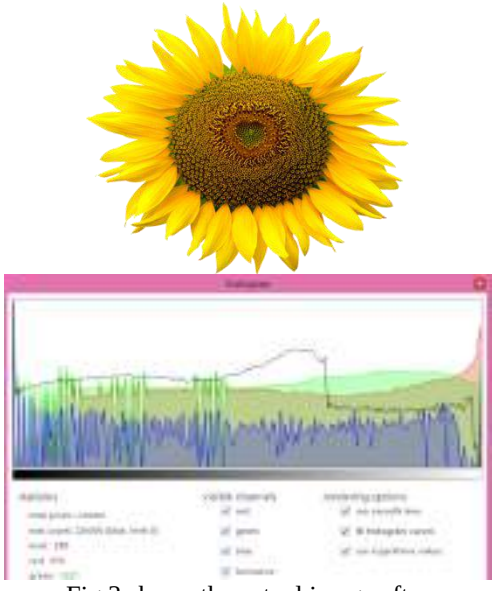


Fig 3 shows the actual image after Decryption

IV. CONCLUSION

In this paper, an ascii formatted message has been hidden into the blocks of image and here the encryption and decryption process plays a major role in this proposed method. In order to obtain the secure Data communication, this proposed work deploys the Steganography using Mainly Nearest neighbor Filling Method, The secret information can be concealed in 0 to N Blocks of image with non overlapping. and decrypt the message with key in linear passion hence the user can be achieved high secure communication.

REFERENCES

[1] J. K. Mandal and Debashis DasColour Image Steganography Based on Pixel Value Differencing in Spatial Domain International Journal of Information Sciences and Techniques (IJIST) Vol.2, No.4, July 2012.

[2] Ajit Singhand UpasanaJauhari , Data Security by Preprocessing the Text with Secret Hiding , Advanced Computing: An International Journal(ACIJ),Vol.3,No.3,May2012DOI:10.5121/acij.2012.330763

[3] Hamdan Lateef Jaheel And Zou Beiji A Novel Approach Of Combining Steganography Algorithms International Journal On Smart Sensing And Intelligent Systems Vol. 8, No. 1, March 2015.

[4] Saiful Islam, Mangat R Modi and Phalguni GuptaEdge-based image steganographyIslam *et al. EURASIP Journal on Information Security* 2014, 2014:8http://jis.eurasipjournals.com/content/2014/1/8

[5] BablooSaha and Shuchi Sharma , Steganographic Techniques of Data Hiding using Digital Images, Received 11 November 2011, online published 23 January 2012 ,Defence Science Journal, Vol. 62, No. 1, January 2012, pp. 11-18, DOI: 10.14429/dsj.62.14362012, DESIDOC

[6] Ross J. Anderson, Fabien A.P. Petit colas, "On the limits of Steganography"

[7] N. Provos, and P. Honeyman, "Hide and Seek: An Introduction to Steganography," IEEE Security and Privacy, 1(3): 32-44,May 2003.

[8] C. Kim, E.-J. Yoon, Y.-S. Hong, and H. 1. Kim, "Secret Sharing Scheme Using Gray Code basedon Steganography," Journal of the Institute of Electronics Engineers of korea, 46(1): 96-102, January 2009.[3] C. C. Thien, and J. C. Lin, "Secret Image Sharing," Computers and Graphics, 26(1):765-770, February 2002.

[9] Graphics, 26[8] Yang, Li "Digital Watermarking". Canada, Ontario. University of Windsor, November 13, 2003.

[10] F. Shih, Digital watermarking and Steganography, fundamentals and techniques. Us SA: CRC Press, 2008. (1):765-770, February 2002.

Splendid Resolution Images Using Contrast Limited Adaptive Histogram Equalization

Sushmitha S.R, Mr. Rajesh K.S B.E
P.G Scholar, Associate Prof
Dept of Computer Science & Engineering,
RajaRajeswari College of Engineering, Bangalore, Karnataka
sushmithasr3@gmail.com, rajeshks_hrr@yahoo.com

ABSTRACT: Medical imaging is a most significant source to find out the presence of assured diseases. Over the years we are using technologies like Computerized Tomography, Positron Emission Tomography and Magnetic Resonance Imaging in medicinal imaging, images are obtained for medical analytical purposes and for providing information about the anatomy, the physiologic and metabolic actions of the volume beneath the skin. It is complicated to get an image at a preferred resolution due to imaging environments, the limitations of physical imaging systems as well as quality-limiting factors such as Noise and Blur. Consequently increasing the image resolution should significantly improve the diagnosis ability for remedial treatment. A counter to this problem is the use of Contrast Limited Adaptive Histogram Equalization and Super Resolution which can be used for processing of such images. This paper details few of the types of medical imaginary, diverse techniques used to perform CLAHE and Super Resolution, how it is used for Medical image processing.
Keywords: Adaptive Histogram Equalization; super resolution; medical images; MRI; CT

1. INTRODUCTION

Medicinal imaging relates a lot of digital image processing systems for improved understanding. Different improvement techniques are available in fiction for refining the quality of medicinal images. The main challenge in this area is that aexplicit process which gives better results for a particular type of presentation may fail in giving good results for another type of application. The objective is to estimation a high-resolution image from a solitary noisy low-resolution image, with the help of aagreed database of high and low-resolution image patch pairs. Denoising and super-resolution in this paper is achieved on each image patch. For individually given input low-resolution patch, its high-resolution form is valued based on outcome a nonnegative sparse linear demonstration of the input patch over the low-resolution patches from the record. The proposed method is expressly useful for the case of noise corrupted and low-resolution image. Investigational results show that the proposed method overtakes other state-of-the-art super-resolution methods while excellentlyeliminating noise.

2. RELATED WORK

[2]EsbenPlenge, enhanced the resolution in magnetic resonance imaging comes at the cost of either lower signal-to-noise ratio, longer possession time or both. This study investigates whether so-called super-resolution reconstruction methods can raise the resolution in the slice selection direction such are a viable alternative to direct high-resolution acquisition in terms of the signal-to-noise ratio and acquisition time trade-offs. The performance of six super-resolution reconstruction methods and direct high-resolution acquisitions was compared with respect to these trade-offs. The methods are based on iterative back-projection, algebraic reconstruction, and regularized least squares. The algorithms were applied to low-resolution data sets within which the images were rotated relative to each other. Quantitative experiments involved a computational phantom and a physical phantom containing structures of known extent. The results show that super-resolution reconstruction can indeed improve the resolution, signal-to-noise ratio and acquisition time trade-offs compared with direct high-resolution acquisition.

[4]This paper explained a unified blind method for multi-image super-resolution, single-image blur deconvolution, and multi-image blur deconvolution of low-resolution images ruined by linear space-invariant blur, aliasing, and additive white Gaussian noise. This approach is based on alternating minimization of a new cost function with respect to the unknown high-resolution image and blurs. The regularization term for the HR image is based upon the Huber-Markov random field model, which is a type of variation integral that exploits the piecewise smooth nature of the HR image. The blur estimation process is supported by an edge-emphasizing smoothing operation. The parameters are rationalized gradually so that the amount of salient edges used for blur estimation increases at all iteration. For better concert the blur estimation is done in the filter area rather than the pixel domain i.e., by means of the gradients of the LR and HR images. Simulation results on both synthetic and real-life images confirm the strength and efficiency of the proposed method.

[7]Jay Patel says Segmentation is a vital role in medical image processing, where clustering technique broadly used in medicinal function mainly for brain tumor recognition in magnetic resonance imaging. Used MRI because of its provide accurate visualize of anatomical structure of tissues. In this paper various clustering methods that have been used for segmentation in MRI are reviewed.

[8]Dinh-Hoan Trinh, Proposed a novel example-based method for denoising and super-resolution of medical images. The objective is to calculate approximately a high-resolution image from a single noisy low-resolution image, with the help of a given record of high and low-resolution

image patch pairs. The problem of finding the nonnegative sparse linear representation is modeled as a nonnegative quadratic programming problem. The planned method is especially useful for the case of noise tainted and low-resolution picture. Investigational outcome show that the proposed system outperforms other state-of-the-art super-resolution methods while effectively removing noise.

According to [10]Pranita Balaji Kanade1 most cells in the body grow and then divide in an orderly way to form new cells as they are needed to keep the body healthy and functioning properly. When cells lose the ability to control their growth, they divide too often and without any sort. The extra cells form a mass of tissue called a tumor. Brain tumors are abnormal and uncontrolled proliferations of cells. Segmentation methods used in biomedical image processing and explores the methods useful for better segmentation. A critical appraisal of the current status of semi automated and automated methods are made for the segmentation of anatomical medical images emphasizing the advantages and disadvantages. In this project we detect the brain tumor & order the stages of the tumor by using testing & training the database. Segmentation for testing purpose is done by spatial FCM used.

3. PROPOSED ARCHITECTURE

In this paper, proposed system consists of two stages, In the first stage the poor qualities of image is processed by preprocessing and patch extraction and in the second stage the output of the first stage is further processed by contrast limited adaptive histogram equalization to improve contrast of images.

3.1 Contrast Limited Adaptive Histogram Equalization

While acting AHE if the area being handled has a fairly small intensity range then the noise in that area gets extra improved. It can also effect some kind of artifacts to seem on those areas. To limit the appearance of such artifacts and noise, alteration of AHE called Contrast Limited AHE can be used. The volume of contrast enhancement for some greatness is directly proportional to the slope of the CDF function at that intensity level. Hence contrast enhancement can be limited by restrictive the slope of the CDF. The slope of CDF at a bin location is unwavering by the height of the histogram for that bin. Consequently if we limit the height of the histogram to a certain level we can border the slope of the CDF and hence the amount of contrast improvement. The only change between regular AHE and CLAHE is that there is one extra step to clip the histogram before the development of its CDF as the mapping function is performed. Hence CLAHE is executed in the same function tiled AHE in ahe.cpp. The sequencer "AHE" takes an extra optional parameter which stipulates the level at which to clip the histogram. By default no clipping is completed.

Following is the summary of the system for this function

1. Compute a grid size based on the full measurement of the image. The smallest grid size is 32 pixels square.
2. If a window size is not indicated chose the grid size as the default window size.
3. Recognize grid points on the image, starting from top-left corner. Each grid point is detached by grid size pixels.
4. For individually grid point compute the histogram of the region around it, having area equal to window size and fixed at the grid point.
5. If a clipping level is stated, clip the histogram computed above to that level and then use the novel histogram to compute the CDF.
6. After computing the mappings for each grid point, recurrence steps 6 to 8 for each pixel in the input image.
7. For individually pixel discover the four closest neighboring grid points that mount that pixel.

Clipping the histogram the situation is not relatively straight forward because the extra after clipping has to be restructured among the other bins, which forcerise the level of the clipped histogram. Later the clipping must be completed at a level lower than the specified clip level so that after restructuring the maximum histogram level is equal to the clip level. To classify the point at which the cutting should be achieved. CLAHE is only active for images which contain relatively similar enhanced noise or artifacts may appear owed to AHE.

3.2 Cubic Spline

Additional piecewise cubic interpolating utility is a cubic spline. The term "spline" states to an device used in enrolling. It is a shrill, stretchy wooden or plastic tool that is agreed through given data points and defines a smooth arch in between. The physical spline diminishes possible energy subject to the interpolation constrictions. The matching mathematical spline must have a uninterrupted second derivative and satisfy the same interpolation constraints. The divisions of a spline are also mentioned to as its knots. The world of splines extends far beyond the basic one-dimensional, cubic, interpolator spline we are labeling here. There are multidimensional, high-order, variable knot, and resembling splines. A valued expository and reference text for both the mathematics and the software is useful guide to Splines.

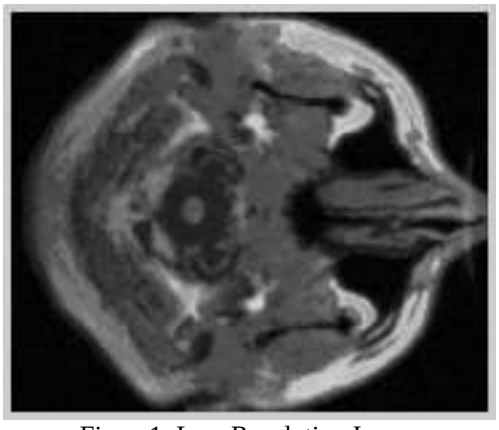


Figure1: Low Resolution Image

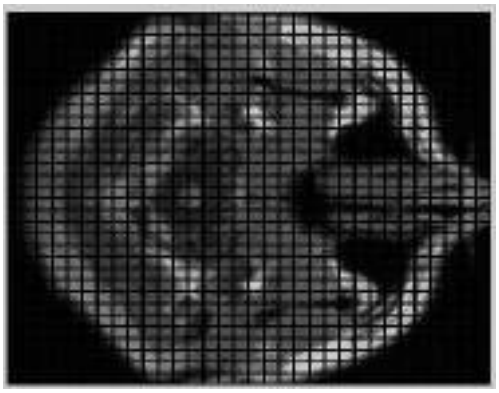


Figure 2: Patch Extraction

Figure 3: Interpolated Image

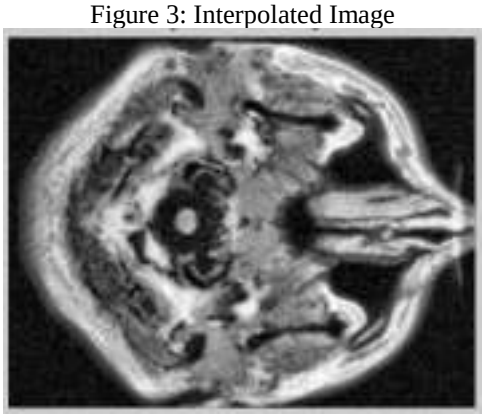


Figure 4: High Resolution Image

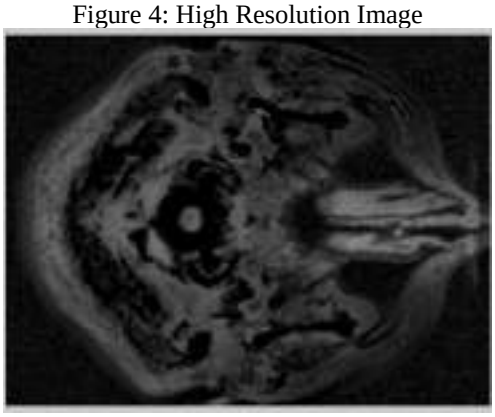


Figure 5: Difference of Inter and High Resolution Image

Here 2 different resolution input is taken, fig1 is the low resolution input image and fig 2 is the patch extracted image of input image. Fig 3 shows the output of interpolated image and fig 5 is the difference of interpolated image and high resolution image.

4. CONCLUSION

Resolution plays an important part in extraction of important information from the images. Better image resolution will help for accurate diagnosis of the ailment and will help in faster rate of treatment to the patients. Have come across the different types of medical images used in Medical image Processing But medical images typically consist of lot of noise and irregularities due to the anatomical structure of the human body and also due to the limitations of the image acquisition device sensors Different methods have been detailed for enhancing the resolution via Super resolution. Dealing with the pre processing part by dynamically enhancing the resolution, de noising the medical images and Later on to apply Contrast Limited Adaptive Histogram Equalization and Super resolution techniques there is tremendous future scope in this method of applying SR techniques for medical image processing.

5. ACKNOWLEDGEMENT

It's my immense pleasure to express my indebtedness to my guide Associate Prof Rajesh K S department of Computer Science and Engineering who guiding me at various stages and I also thank the principal and management of the RajaRajeswari College of Engineering, Bangalore.

REFERENCES

[1]. Thirimachos Bourlai, Member, IEEE, Arnos, Senior Member, IEEE, and Anil K. Jain, Fellow, IEEE," Restoring Degraded Face Images: A Case Study in Matching Faxed, Printed, and Scanned Photos", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 6, NO. 2, JUNE 2011.

[2]. Esben Plenge,1,2* Dirk H. J. Poot,1,2 Monique Bernsen,3 Gyula Kotek,3 Gavin Houston,3 Piotr Wielopolski,3 Louise van der Weerd,4,5 Wiro J. Niessen,1,2,6 and Erik Meijering1,2," Super-Resolution Methods in MRI: Can They Improve the Trade-Off Between Resolution, Signal-to-Noise Ratio, and Acquisition Time?", © 2012 Wiley Periodicals, Inc.

[3]. S. Chaudhari, Department of Electrical Engineering, Indian Institute of Technology – Bombay, Mumbai, India 400 076, "Super Resolution Imaging". New York, Boston, Dordrecht, London, Moscow, Kluwer Academic Publishers.

[4]. Esmaeil Faramarzi, Member, IEEE, Dinesh Rajan, Senior Member, IEEE, and Marc P. Christensen, Senior Member, IEEE," Unified Blind Method for Multi-Image Super-Resolution and Single/Multi-Image Blur Deconvolution", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 6, JUNE 2013.

[5]. F. Rousseau, D. Gounot and C. Studholme, "On High-Resolution Image Estimation using Low-Resolution Brain MRI", 35th Annual International Conference of the IEEE EMBS, Osaka, Japan, 3 - 7 July,2013.

[6]. D. Trinh, M. Luong, F. Dibos, J. M. Rocchisani, C. D. Pham and T. Q. Nguyen, "Novel Example-Based Method for Super-Resolution and Denoising of Medical

Images", IEEE Transactions on Image processing, Vol. 23, No. 4, April 2014.

[7]. Jay Patel1 and Kaushal Doshi M.Tech Student, EC Department, Gujarat Technology University, Kalawad Road, Rajkot-360001, India. Assistant Professor, EC Department, MEFGI Collage, Rajkot Kalawad Road, Rajkot-360001, India.," A Study of Segmentation Methods for Detection of Tumor in Brain MRI", Advance in Electronic and Electric Engineering.ISSN 2231-1297, Volume 4, Number 3 (2014), pp. 279-284© Research India Publications.

[8]. Dinh-Hoan Trinh, Member, IEEE, Marie Luong, Member, IEEE, Françoise Dibos, Jean-Marie Rocchisani, Canh-Duong Pham, and Truong Q. Nguyen, Fellow, IEEE," Novel Example-Based Method for Super-Resolution and Denoising of Medical Images", 1057-7149 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

[9]. H. Okuhata, R. Imai, M. Ise, R. Y. Omaki1, H. Nakamura, S. Hara and I. Shirakawa, "Implementation of Dynamic-Range Enhancement and Super-Resolution Algorithms for Medical Image Processing" in 2014 IEEE International Conference on Consumer Electronics (ICCE).

[10]. Pranita Balaji Kanade1, Prof. P.P. GumasteDepartment of E&TC Engineering, JSPMs, Jayawantrao Sawant College of Engg, Hadapsar, Pune, India,"Brain Tumor Detection Using MRI Images", INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING Vol. 3, Issue 2, February 2015.

An Interactive Smart Glass

Srinivas M[1], Gangaram D[2], Varun Reddy R[3], AtulKewat K[4]
UG students, Dept. of CSE, RRCE, Bengaluru
Dr. UshaSakthivel[5], HOD& Prof.,Dept. of CSE, RRCE, Bengaluru

ABSTRACT-Our research goal is to make the world smarter in day to day life. Nowadays, everything in the world is getting smarter i.e. for ex: smart phone, smart TV etc. Our project is mainly based on this concept, nothing but an interactive smart glass. It implies several constraints. First, conversion of image to text and produces output by voice. Second, Object recognition by using IR rays. Third, Enhancement of old documentand detect the content present in it. Fourth, Recognition of ancient or hand written text content in the document. Fifth, Navigating the location to the user by GPS mode and displaying it on the screen. All the above constraint can be fulfilled by using Text detection, Enhancement process, Object recognition method embedded in normal glass with android functionality. And Hence, we all this glass as an Interactive Smart Glass where User can interact with it by providing input to it.

Keywords:Image Segmentation, Optical character recognition(OCR), Text detection, Object recognition, Thresholding.

I. INTRODUCTION

Today, there are about 2-3% people of world population are blind and low vision paired people. We all know that blind people have own script language known as Braille language, which is slightly complex to learn. If a person has low vision or who loses his eye sight due to some accident might feel bit difficult to learn the Braille language. So in order to overcome difficulties faced by thisspecial people we have researched and proposed a model known as an Interactive Smart Glass, where user can read the document by capturing the image and detecting the text content in it and output is produced through speech to the user. Secondly,User can also recognize the object present in front of camera(which is present in smart glass) by IR rays for about 5m distance. This smart glass can be used by even normal people, as it helps in navigating the location to the user through GPS and display it on the glass screen.And it helps in enhancing the old document and detecting the content present in it. And also detect the ancient or handwritten text present in the document through text detection process and produces the output as speech to the user. For processing all this we are using android function and RAM for this. And we also use image segmentation, object recognition, text detection method which will be embedded in android environment.

II. RELATED TECHNOLOGY

Previously, for converting the digital image into text and speech, we need to perform it by using MATLAB concept. This concept is based on embedded system term which will run on the computer system. Even for converting the text to an image we were using Text to Speech synthesis. The above methods take more processing time, so in order to overcome this we are proposing Image to Text and Speech conversion, which will be developed and embedded in android platform which user friendly.

III. MECHANISM

In order to implement the above proposed work, we need to follow some methods. The methods are as follow:

3.1 Document Enhancement

Here, we concentrate on enhancing the old document and detecting the content present in it. We capture or upload the image which we need to enhance and the image is being processed for enhancing. After a few second, image will be enhanced, which will be better version of previous. After the enhancement process is completed, Text detection and image segmentation will be applied to it. And then content present is detected in it by the method and give the output through voice to the user. During the enhancement process it will undergo following step: De-bluring of the image, Interpolation method, and applying adaptive threshold to that image. All these methods are executed in android application which is developed in android environment

.

Fig3.1: Document Enhancement

3.2 Object Recognition

Here, we look on the recognition of the object by using Infrared rays. The camera present in the smart glass will capture or scan the scene present infront of it. Through IR rays it detects the object present in it. Infrared rays can

travel up to some extent for about 5m distance. It will recognize the object present and produce output through voice to the user. There are few which will undergo during object recognition. First, Alignment method. secondly, invariants properties will be applied. And finally, Decomposition of parts will be applied.All the methods are done in android environment, where we use RAM of 1GB and Quad-core processor for it.

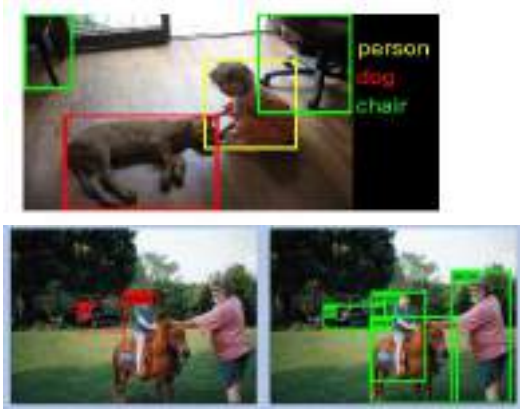


Fig3.2 (a): Object Recognition

The above picture depicts the object recognition. Here, an image is captured and subjected to object recognition which is being recognized by using IR rays. And IR rays detect the object present in front of the camera.

Fig3.2 (b): The device used for object recognition using IR rays.

Fig3.2(c): Flow diagram of Object recognition.

3.3 Navigation

Navigation means showing the direction. Here, the concept of navigation is also applied in our model in order to navigate the place by using WiFi and GPS mode. In this feature if we give input to the smart glass through voice, it will navigate us to the location. The navigation of the place will be displayed on the screen of the glass and voice output will also be produced to the user. This feature will use Global positioning system for navigating the location. And it will be helpful for all people.

Fig3.3: Navigation

3.4 Text to Speech Conversion

Here, we mainly concentrate on conversion of text to speech. Firstly, we capture the image through a camera, and then image is being pre-processed where it undergoes, following method: Feature extraction, image segmentation, edge detection, optical character recognition and Text synthesis to speech. After undergoing all this process an image containing text will be synthesized and produce the output as speech to the user.By this it will be able to convert the image to text and speech.And this feature will be more helpful for low pair vision and Blind people to read the books or a document.

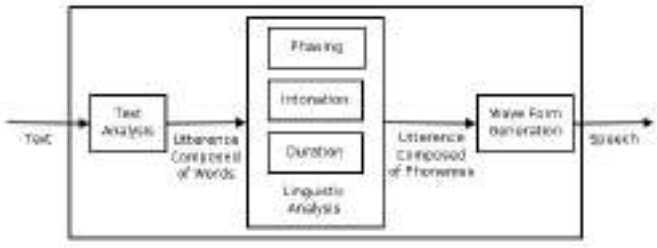


Fig 3.4: Text to Speech Synthesis

In the above fig Text to Speech synthesis flow is shown. Here, Text will be taken as input and it is analyzed and then it is synthesized to speech.

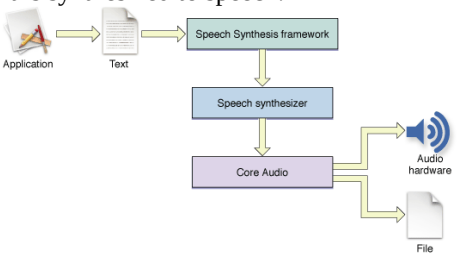


Fig 3.4(a): Text to Speech flow

3.5Text Detection

In this we mainly focus on Text Detection concept. This method is also same as that of object recognition but here it detects only text present in the image.

Fig 3.5(a): Text Detection

As you see in the above picture, the smart glass captures an image and detects the text present in the image. Recognition of the text is done by preprocessing of the image. Preprocessing stage may involve image segmentation, edge detection and text detection methods. All the methods are developed as an android application on android platform and embedded in the smart glass.

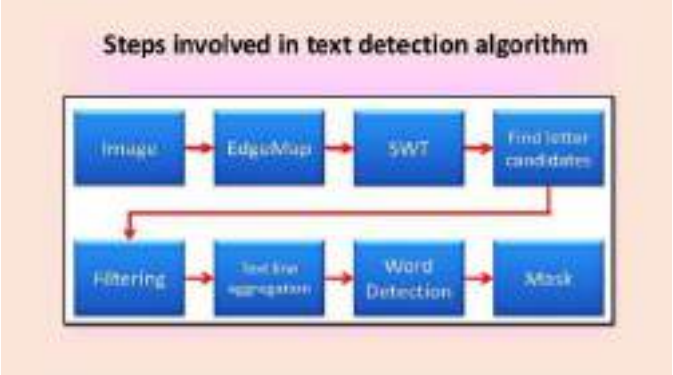


Fig 3.5(b): Steps involved in Text Detection Algorithm

As the image only depicts the flow of the algorithm, where image captured is preprocessed, edge detection method is applied to that image and further filters are applied to that image inorder to detect the text present in the image.

IV. REAL TIME REQUIREMENTS

In order to build this proposed work, user feedback plays a vital role in it. Lot of people are interested in android, so for this purpose we are using android functionality to build this model.We also require a Quad core processor with 1GB of RAM for speed processing. If the size of RAM is more, then performance of a smart glass will be higher. We also use High Definition of camera and IR Rays for processing some methods in it.

V. FUTURE ENHANCEMENT

Efforts are being made for further enhancing of this device. Our new ideas towards enhancing the device. Firstly,3D object modelling view where you can view the object in 3D view and interact with it by using hand gestures. Secondly, to view the 2D image into 3D image in High Definition mode.

VI. CONCLUSIONS

We have compiled research trends in an Interactive Smart Glass using Google scholar's searched tool citation found exciting. Here image segmentation, object recognition and text detection feedback have received a lot of attention, and hence further proceeding towards future enhancement of the device which will be useful to society.

VII. ACKNOWLEDGEMENT

The documentation and implementation of this paper would not be succeeded without the kind support from Dr. R BalaKrishna, principal, RajaRajeshwari College of Engineering, Bengaluru, who always gives us valuable advice and kind assistance to complete this paper. We are grateful to Dr.UshaSakthivel, Professor and HOD, Department Of Computer Science and Engineering, RajaRajeswari College Of Engineering,Bengaluru, for giving us the great knowledge.

Finally, we would like to thank our guide and parents who constantly supported us since the beginning till the end of the paper.

REFERENCES

[1] L. von Ahn and L. Dabbish, "Labeling Images with a Computer Game," *Proc. ACM CHI*, 2004.

[2] Airliners.net, http://www.airliners.net.

[3] J. Amores, N. Sebe, P. Radeva, T. Gevers, and A. Smeulders, "Boosting Contextual Information in Content-Based Image Retrieval," *Proc. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*, 2004.

[4] J. Assfalg, A. Del Bimbo, and P. Pala, "Three-Dimensional
Interfaces for Querying by Example in Content-Based Image Retrieval," *IEEE Trans. Visualization and Computer Graphics*, 8(4):305–318, 2002.

[5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching Words and Pictures," *Journal of Machine Learning Research*, 3:1107-1135, 2003.

[6] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate
Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.

[7] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and
Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[8] D. M. Blei and M. I. Jordan, "Modeling Annotated Data,"
Proc. ACM Conference on Research and Development in Information Retrieval, 2003

[9] C. Bohm, S. Berchtold, and D. A. Keim, "Searching in High-Dimensional SpacesIndex Structures for Improving the Performance of Multimedia Databases," *ACM Computing*

Surveys, 33(3):322-373, 2001.
[10] J. Carballido-Gamio, S. Belongie, and S. Majumdar "Normalized Cuts in 3-D for Spinal MRI Segmentation," *IEEE Trans. Medical Imaging*, 23(1):36–44, 2004.
[11] G. Carneiro and N. Vasconcelos, "Minimum Bayes Error
Features for Visual Recognition by Sequential Feature Selection and Extraction," *Proc. Canadian Conference on Computer and Robot Vision*, 2005.
[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-maximization and Its Application to Image Querying," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(8):1026-1038, 2002.
[13] A. Chalechale, G. Naghdy, and A. Mertins, "Sketch-Based
Image Matching Using Angular Partitioning," *IEEE Trans. Systems, Man, and Cybernetics*, 35(1):28–41, 2005.
[14] E. Y. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA: Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines," *IEEE Trans. Circuits and Systems for Video Technology*, 13(1):26–38, 2003.
[15] C.-c. Chen, H. Wactlar, J. Z. Wang, and K. Kiernan, "Digital Imagery for Significant Cultural and Historical Materials - An Emerging Research Field Bridging People, Culture, and Technologies," *International Journal on Digital Libraries*, 5(4):275–286, 2005.
[16] J. Chen, T.N. Pappas, A. Mojsilovic, and B. Rogowitz, "Adaptive image segmentation based on color and texture," *Proc. IEEE International Conference on Image Processing*, 2002.
[17] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature
Matching Approach to Content-Based Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):252–1267, 2002.
[18] Y. Chen and J. Z. Wang, "Image Categorization by Learning and Reasoning with Regions," *Journal of Machine Learning Research*, 5:913-939, 2004.
[19] Y. Chen, J. Z. Wang, and R. Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," *IEEE Trans. Image Processing*, 14(8):1187–1201,
2005.
[20] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5), 603–619, 2002.
[21] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments," *IEEE Trans. Image Processing*, 9(1):20–37, 2000.
[22] A. Csillaghy, H. Hinterberger, and A.O. Benz, "Content Based Image Retrieval in Astronomy," *Information Retrieval*, 3(3):229–241, 2000.
[23] S. J. Cunningham, D. Bainbridge, and M. Masoodian, "How People Describe Their Image Information Needs: A Grounded Theory Analysis of Visual Arts Queries," *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2004.
[24] C. Dagli and T. S. Huang, "A Framework for Grid-Based
Image Retrieval," *Proc. IEEE International Conference on Pattern recognition*, 2004.
[25] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and
H. Shin, "An Efficient Color Representation for Image Retrieval," *IEEE Trans. Image Processing*, 10(1):140–147, 2001.
[26] M. N. Do and M. Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," *IEEE Trans. Image Processing*, 11(2):146–158, 2002.
[27] A. Dong and B. Bhanu, "Active Concept Learning for Image Retrieval in Dynamic Databases," *Proc. IEEE International Conference on Computer Vision*, 2003.
[28] Y. Du and J. Z. Wang "A Scalable Integrated Region-Based Image Retrieval System," *Proc. IEEE International Conference on Image Processing*, 2001.
[29] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Seventh European Conference on Computer Vision, pages 97–112, 2002.
[30] Global Memory Net, http://www.memorynet.org.
[31] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, andR. Jain, "Content-Based Image Retrieval at the End of theEarly Years," *IEEE Trans. Pattern Analysis and MachineIntelligence*, 22(12):1349–1380, 2000.

Identifying suitable enhancement technique for thermal and non thermal diabetic foot images

Dr.Punal M Arabi, Gayatri Joshi, Tejaswi Bhat , Harsha L Singh
Department of Biomedical Engineering, ACSCE, Bangalore-74
arabi.punal@gmail.com, gayitrijoshi@gmail.com, tbhat1995@gmail.com, puthul13@gmail.com

ABSTRACT : Image processing plays a vital role in disease diagnosis, treatment planning and monitoring. The Images acquired by various modalities like MRI(Magnetic resonance imaging), ultrasound , x-ray, PET(Positron emission tomography) are all to be processed to extract information that are crucial for diagnosis and treatment. The noise removal and contrast enhancement are the key steps in preprocessing of an image which make the image suitable for further processing. This paper probes the suitability of contrast stretching and histogram equalization methods of image enhancement techniques for thermal and non thermal diabetic foot images. Thermal and non thermal Images of diabetic and normal foot are taken filtered and enhanced. The SNR(signal to noise ratio) is then calculated to find the suitable method of image enhancement for thermographic, non thermographic diabetic foot images. The results obtained show that contrast stretching is giving better SNR(signal to noise ratio) for thermographic diabetic foot image where as histogram equalization is superior in performance in case of non thermo graphic diabetic foot image enhancement.

Key words: Diabetic foot-,Thermographic,Non thermographic ,Contrast Limited Adaptive Histogram Equalization (CLAHE), Contrast stretching ,Signal to noise ratio(SNR).

I. INTRODUCTION

The difference in visual properties that makes an object (or its representation in an image) distinguishable from other object is known as contrast. In the visual perception of real world, contrast is determined by the difference in the colour and brightness of the object. The principal objective of enhancement is to process an image to achieve reliable results. Histogram is defined as the graphical distribution of pixels over the luminous intensity. In statistics, a histogram is a graphical representation of the distribution of data. It is the estimatation of the probability distribution of a continuous variable. The histogram equalization is an approach to enhance a given image. Contrast stretching increases the dynamic range of the gray levels in the image.Thermography is a procees of printing or imaging method that uses heat to create an image. The imaging method thermographic imaging captures infrared energy emitted by the subject. Medical Thermography (digital infrared thermal imaging - DITI) is the method for early pre-clinical diagnosis and control during treatment of homeostatic imbalances.

Active thermography is an effective tool for non-destructive evaluation. It includes modulated (lock-in) or pulsed thermography, Thermography is a non-invasive, non-contact tool that uses the heat from our body to aid in making diagnosis of a subject under investigation and is completely safe and uses no radiation.

II Literature survey

M.Aarthy.etal[1],described three different techniques of expansion namely dynamic range expansion, linear contrast expansion and symmetric range expansion.Each of these has their specific strength and weakness. For colored images linear contrast expansion is used. These all methods help in easy study of histograms and helps in image enhancement. Salem Saleh Al-amri.etal[2**],** attempted to undertake the two types of the contrast enhancement techniques, linear contrast techniques and non-linear contrast techniques. In linear contrast techniques applied three methods, Max-Min contrast method, Percentage contrast method and Piecewise contrast technique. Non-linear contrast techniques applied four contrast methods, Histogram equalization method, Adaptive histogram equalization method,Homomorphic Filter method and Unsharpe Mask. Raman Maini.etal[3], provided an overview of underlying concepts, along with algorithms commonly used for image enhancement. They focused on spatial domain techniques for image enhancement, with particular reference to point processing methods and histogram processing. Almar Klein[4],addressed the problem of aligning the images of feet taken at different instances in time. They proposed to use keypoints to find the geometric deformation between two photo', then have a set of landmarks for each image.By finding the corresponding landmarks (i.e. matching the keypoints),They know the deformation in a sparse set of points.Using these points, calculated a B-spline deformation model, which allows us to describe the deformation in every point in the image. Sanjeev Kumar.etal[5], proposed with contrast enhancement of X-

Ray images and presents here a new approach for contrast enhancement. Comparative analysis of proposed technique against the existing major contrast enhancement techniques has been performed. Here, the results of each technique were illustrated for various backgrounds, majority of them in poor lighting condition. . S. Srinivasan[6], described a real time contrast enhancement technique for digital video applications.This method called ACE is based on a modified histogram equalization procedure that adapts to the input video statistics. The method decides whether to increase dynamic range or to light up dark regions of the image. Mrs. Pallavi Mahajan.etal[7],proposed the state of the art of image processing techniques to detect the thyroid gland disease non- traumatically using Thermograph. Thermal Imaging is a technology that creates analyzes images by detecting the heat radiating from an object. They have proposed a system to detect the thyroid gland disease using thermograph.C.Ibarra-Castanedo.etal [8],proposed raw images are not often appropriate since most will be missed. In some other cases, what is needed is a quantitative analysis such as for defect detection and characterization. Presented was made of various methods of data analysis required either at preprocessing and/or processing of thermo graphhic images.

In this work tells about the novel method of identifying the enhanced methods for thermal and non thermal diabetic foot images.

III METHODOLOGY

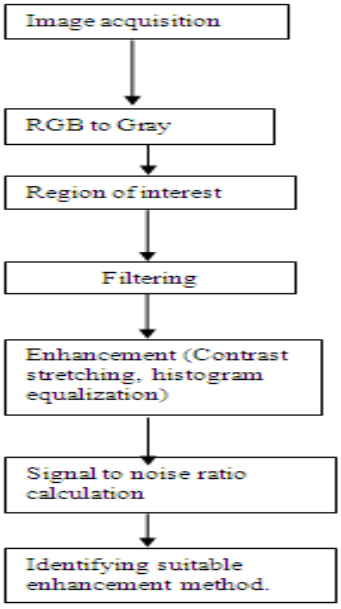


Fig1: Flow diagram

Thermal and non thermal diabetic foot images are taken for analysis. The color images are converted to gray. The region of interest (ROI) is then selected. The ROI must have at least 600 to 1200 pixels to get reliable results. Different enhancement techniques like contrast stretching, histogram equalization are applied for both thermo graphic and non thermo graphic foot images. The ROI of these images are taken and the signals to noise ratio values are then calculated. The calculated signal to noise ratio values of the filtered images by different enhancement techniques have been compared to analyze the performance of the contrast stretching and histogram equalization for thermographic and non thermo graphic diabetic foot images.

The formulae of Contrast stretching Histogram equalization and SNR are as follows:

Histogram: $n=\sum_{i=1}^{k} m_i$……..(1)

n be the total number of observations and *k* be the total number of bins, the histogram m_i

Contrast stretching:

New_value=[(Old_value-Min_value)/(Max_value-Min_value)]*255……..(2)

Signal to noise ratio:

$SNR=\frac{P_{signal}}{P_{noise}}$ ……..(3)

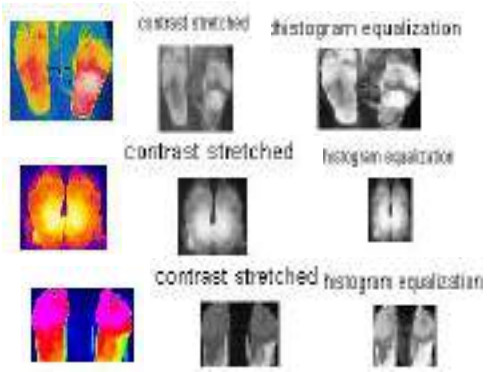
where *P* is average power. Both signal and noise power must be measured at the same or equivalent points in a system, and within the same system bandwidth.

IV Results

S.NO	Images	SNR(signal to noise ratio)	
		Contrast stretching	Histogram equalization
1	Thermo graphic diabetic foot1	14.0772	10.6510
2	Thermo graphic diabetic foot2	13.45	9.34
3	Thermo graphic diabetic foot3	15.56	11.78
4	Non thermal diabetic foot1	10.56	15.654
5	Non thermal diabetic foot2	9.80	13.67
6	Non thermal diabetic foot3	11.78	14.56

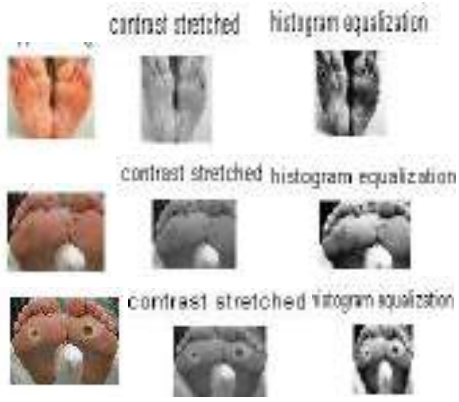
Table1:Signal to noise ratio of contrast stretching and histogram equalization

Fig2(i): Thermal diabetic foot (ii) contrast stretched diabetic foot image (iii)histogram equalization of diabetic foot image



2 (i) 2(ii) 2(iii)

Fig3(i): Non thermal diabetic foot image (ii) contrast stretched diabetic foot image (iii) histogram equalization of diabetic foot image



3(i) 3(ii) 3(iii)

Figure 2(i) shows the thermal diabetic foot image,Fig2(ii) shows the contrast stretched thermographic diabetic foot image and fig 2(iii) shows the histogram equalization of thermal diabetic image . Figure 3(i) shows the non thermal diabetic foot image,Fig3(ii) shows the contrast stretched non thermo graphic diabetic foot image and fig 3(iii) shows the histogram equalization of non thermal diabetic image.

Table1 shows the Signal to noise ratio values of the thermal and non thermal diabetic foot images using different enhancement techniques.

V. Discussion

Three set of thermal and non thermo graphic diabetic foot images are taken for experimentation. A region of interest is identified in each sample and various enhancement methods are used to enhance this region of interest. Signal to noise ratio values are found for the enhanced images.

The results tabulated in Table-1 show that the contrast stretching method of thermo graphic diabetic foot is having better signal to noise ratio value compare to other enhancement method i.e: histogram equalization is having lower signal to noise ratio value. For the non thermo graphic diabetic foot the signal to noise ratio for two enhancement methods are following reverse pattern i.e: contrast stretching is less and for histogram equalization it is higher. From the tabulated signal to noise ratio results it is seen that of two enhanced methods considered here the contrast stretching results in higher signal to noise ratio while enhancing the thermo graphic diabetic foot image and for non thermo graphic images histogram equalisation gives higher SNR. Hence contrast stretching is superior to histogram equalisation enhancement method in case of thermo graphic foot image processing where as histogram equalisation enhancement method performs better than contrast stretching in the case of non thermo graphic diabetic foot images.

VI. Conclusion

The proposed method is tested on three sets of thermo graphic and non thermo graphic diabetic foot images. The results obtained show that the thermo graphic diabetic foot is having higher signal to noise ratio for contrast stretched enhancement method compared to the histogram equalization method. It is found that the contrast stretched enhancement method for thermo graphic diabetic foot image properly enhanced. Whereas histogram equalization method is having signal to noise ratio is lesser compared to contrast stretched method. In non thermo graphic diabetic foot image the signal to noise ratio values for both the enhancement methods are following reverse pattern compared to thermo graphic diabetic foot image i.e., histogram equalisation enhancement method performs better than contrast stretching in the case of non thermo graphic diabetic foot images. Hence contrast stretching is superior to histogram equalization enhancement method in the case of thermo graphic foot image processing where as histogram equalization enhancement method performs better than contrast stretching in the case of non thermo graphic diabetic foot images.

However more samples of thermo graphic and non thermo graphic diabetic foot images are to be tested to know about the accuracy of the proposed method.

Acknowledgment

The authors thank the Management and Principal of ACS College of engineering, Mysore road, Bangalore for permitting and supporting us to carrying out this research work.

References

[1]. M.Aarthy, P.Sumathy," A Comparison Of Histogram Equalization Method And Histogram Expansion",

International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 3, pg. 25-34 ISSN: 2321-8363, March- 2014

[2].SalemSalehAl-,Dr.N.V.Kalyankar,Dr.S.D.Khamitkar, "Linear and Non-linear Contrast Enhancement Image", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.2, February 2010

[3].Raman Maini And Himanshu Aggarwal." A Comprehensive Review Of Image Enhancement Techniques", Journal Of Computing, Volume 2, Issue, Issn 2151-96173, March 2010

[4]. Almar Klein, F. Van Der Heijden, C.H. Slump," Alignment Of Diabetic Feet Images", Proceedings Of Sps-Darts 2007 (The 2007 The Third Annual Ieee Benelux/Dsp Valley Signal Processing Symposium

[5].Sanjeev Kumar Dr.VijayDhir sourabhMehra,"Analysis & Implementation of Contrast Enhancement Techniques Using Medical Image",International Journal of Advanced Research in Computer Science and Software Engineering,Volume 4, Issue 5,ISSN: 2277 128X, , May 2014

[6]. S. Srinivasan, N. Balram," Adaptive Contrast Enhancement Using Local Region Stretching", Proc.of ASID'06, 8-12 Oct, New Delhi.

[7]. Mrs. Pallavi Mahajan, Mrs. Swati Madhe," Morphological Feature Extraction of Thermal Images for Thyroid Detection", International Journal of Electronics Communication and Computer Engineering Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

[8]. C. Ibarra-Castanedo, D. Gonz alez, M. Klein, M. Pilla, S. Vallerand, X. Maldague ," Infrared image processing and data analysis", 1350-4495/$,Published by ElsevierB.V. doi:10.1016/j.infrared.2004.03.011

9. N.R.Mokhtar, N. H. Harun, M.Y.Mashor, H.Roseline , N.Mustafa, R.Adollah ,H. Adilah, N.F.Mohd Nasir,"Proceedings of the World Congress on Engineering ",Vol 1 WCE 2009, July 1 - 3, London, U.K,2009.

10. Y.Y. Wang, T. Hua, B. Zhu, Q. Li, W.J. Yi, X.M. Tao, "Novel fabric pressure sensors: design, fabrication and characterization", Smart Materials and Structures, vol.20, 065015, 2011.

11. L. Shu, K.Y. Mai, X.M. Tao, W.C.Wong, K.F. Lee, S. L. Yip, W. H.Shum, W.L. Chan, C. P.Yuen, Y. Li, S.F. Yau, "Novel Intelligent Footwear System: Dynamic Plantar Pressure monitoring of Diabetic Patients during Daily Activities", The 9th Asia Pacific Conference on Diabetic Limb Problems, November 23-25, 2012.

A Framework for Anomaly Event Detection by Analysing the Video Sequences

Puttegowda D, U N Sinha
Research Scholar, National Aerospace Laboratories , Bangalore
Former Distinguished Scientist, CSIR-CMMACS, Bangalore
puttegowda.77@gmail.com

ABSTRACT : Currently, it is very essential to monitor activities in Video surveillance applications both in private and public environments. In this context, our paper presents a novel framework to detect the normal or abnormal event situations by analyzing the pixel-wise motion context using block-based approaches. First motion estimation techniques are applied to characterize the events at the pixel level. Optical flow is used to extract information such as density and velocity of motion. The two different proposed approaches identifies abnormal motion variations in regions of motion activity based on the entropy of Discrete Wavelet Transform and Discrete Cosine Transform coefficients. The successful results of the detection of normal or abnormal events on different datasets are reported.

Keywords – Discrete Cosine Transform, Discrete Wavelet Transform, event detection, optical flow, Video surveillance.

I. INTRODUCTION

To ensue security at airports, banks, and institutions, Video surveillance has been a key component [1]. The main intent of visual surveillance is not only to use cameras instead of human eyes, but also the task of surveillance is carried out automatically by analyzing the video sequences. For automatic dynamic scene analysis, the challenging task for anomaly detection especially given a scene consisting of activities of multiple objects [2]. Anomaly detection techniques can be broadly classified into two broad families of approaches like pattern recognition-based and machine-learning-based methods. In pattern recognition methods, prior knowing the type of abnormal activity or object is must. But, the recognition methods require a list of objects or behavior patterns that are anomalous. Unfortunately, suspicious activities cannot be known in advance is not always possible. In machine learning based approach, by exhibiting regular activity in a video sequence, the behavior of the object is detected and then flag moving objects whose behavior deviates from normal behavior [3].

Recently, many authors have proposed learning methods based on characteristics other than motion paths to address these limitations [8]. In such a case, there is no need for object tracking, instead, features at pixel-level are considered. The main idea here, instead of tracking subjects one by one, general motion context are analyzed. We proposed a general framework based approach, where features like velocity at pixel-level are directly extracted from motion. This will lead to an image that expresses the motion in the scene. Then we analyze the information content of that image in the frequency domain by computing the entropy of the involved DWT(Discrete Wavelet Transform) or DCT(Discrete Cosine Transform) coefficients. After successfully analyzing motion in each frame, we should understand the behavior of the objects. To understand the Behavior of the objects, it involves the analysis and recognition of motion patterns, the description of actions and interactions at high level. We compare the entropies for each block to the median averaged values over time to classify events into normal and abnormal based on motion features.

The paper consisting of following sections. Section II describes the two proposed approaches for abnormal event detection including motion estimation, measuring entropy and then detecting abnormal events using DWT and DCT techniques. Section III gives experimental results and followed by Section IV conclusion and References.

II. RELATED WORKS

In the papers [4][5], a general pipeline-based framework is implemented. After detecting the moving objects initially, then by considering certain number of frames, objects are classified and tracked and finally, the activity of normal behavior of objects from the abnormal behavior are distinguished by using resulting path. Many different types of applications has proven successful different tracking-based methods, but they suffer from fundamental restrictions . First, implementing such a pipeline method can result in a delicate architecture which may go through the subsequent processing stages with an error. Secondly, tracking multiple objects at the same time requires complex algorithms [6][7]. Therefore, multi-object tracking is not always easy and efficient in crowded areas where objects occluded each other regularly fully or partially. This task is spatially hard where quality and color information can be poor in surveillance videos.

In the paper[14], temporal and spatial anomalies is proposed to detection and localization of anomalous

behaviors in crowded scenes. The events in Surveillance video are extremely tedious to monitor which has very low probability. To enumerate the set of anomalies in crowds scenes that are possible in a given surveillance scenario is also infeasible. In this context, our paper presents a novel framework to detect the normal or abnormal event situations by analyzing the pixel-wise motion context using block-based approaches.

III .PROPOSED APPROACHES

A. Abnormal event detection using DWT

Our abnormal event detection is based on motion features extracted with a motion estimation technique. Motion estimation in image sequences aims at detecting regions corresponding to moving objects. For tracking and behavior analysis in the later stage, the features of moving object is detected. For noisy and coarse data, the computation of optical flow is not accurate. To deal with this, we use optical flow at each frame using the Lucas-Kanade algorithm [9]. To detect the moving regions in an image sequence relating each image to the next is used in Optical-flow-based motion estimation is considered by the characteristic of flow vectors of moving objects over time. The perceptible displacement of each pixel from image to image is represented in each vector [10]. The result of optical flow at both vertical and horizontal direction is the value of displacement of each pixel. To obtain a motion magnitude vector, we combine this displacement. To progression these motion vectors, we replace with pixel values for the estimated motion and each frame is divided into blocks. We anticipate that during abnormal events the motion patterns and the energy of the images containing motion vectors change compared to normal behavior. Here we use DWT to find the coefficients. We apply a DWT to each block, as the DWT provides a compact representation of the signals energy. Then we compute the entropy of the DWT coefficients to measure the information content of the DWT coefficients.

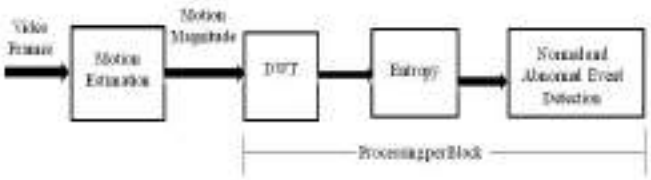


Fig. 1: Proposed approach block diagram

The Fig.1 shows how actually the design of our proposed work is, where the processing is done block by block. Firstly, the video will be given as a input, that video will be divided into number of frames, these video frames are given as input to the Motion Estimation block.

The output of the Motion Estimation block will be the motion magnitude, this motion magnitude will be the input to the Discrete Wavelet Transform (DWT) block. In DWT block only level one is used. The four parts of the DWT are LL, LH, HL, HH. Out of these four parts, only three parts like LH, HL and HH are used for object detection. In the next block the high pass outputs i.e., HH, HL and LH parts are combined together which results in a form in which most of the image pixels becomes dark except the outline of the object which is moving in the input video. The output of the Discrete Wavelet Transform (DWT) block will be of coefficient values and these values will be the input to the Entropy block.

The output of the Entropy block will decide whether the event present in the given video data set is Normal event or Abnormal event. This is the output of this below block diagram that is output of our proposed system. Processing of this below block diagram will be done per block.

B. Abnormal event detection using DCT

The second method for abnormal event detection using DCT. Using the Motion estimation of image sequences are used to detecting regions corresponding to moving objects. For tracking and behavior analysis in the later stage, the features of moving object is detected. . For noisy and coarse data, the computation of optical flow is not accurate. To deal with this, we use optical flow at each frame using the Lucas-Kanade algorithm. To detect the moving regions in an image sequence relating each image to the next is used in Optical-flow-based motion estimation is considered by the characteristic of flow vectors of moving objects over time. The perceptible displacement of each pixel from image to image is represented in each vector. The result of optical flow at both vertical and horizontal direction is the value of displacement of each pixel. To obtain a motion magnitude vector, we combine this displacement. To progression these motion vectors, we replace with pixel values for the estimated motion and each frame is divided into blocks. We anticipate that during abnormal events the motion patterns and the energy of the images containing motion vectors change compared to normal behavior. To find the coefficients, here we used Discrete Cosine Transform. DCT provides a compact representation of the signals energy when we apply DCT to each block. Then we compute the entropy of the DCT coefficients to measure the information content of the DCT coefficients.

$$T = (u,v)\alpha_u \alpha_u \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)P.Q \quad (1)$$

Where

$$P = \cos \frac{(2x+1)u\pi}{2M}$$

and

$$Q = \cos \frac{(2y+1)v\pi}{2N}$$

$$\alpha_u = \alpha_v = \begin{cases} \sqrt{\frac{1}{M}} & u = v = 0 \\ \sqrt{\frac{2}{N}} & u \neq v \neq 0 \end{cases}$$

f(x,y) = input image with size M x N, where M=row and N=Column

The entropy is defined as [12]:

$$E = -\sum_{i=1}^{N} p \log p \qquad (2)$$

where N is the size of image and at a certain pixel location, probability of the motion intensity value is stored in p. We compare the entropy value with thresholds which we learn per block in the beginning of the video sequence to decide whether event is normal event or abnormal event. By considering the first 100 frames of the video, the threshold value is estimated based on a median value of the entropies. When the value of the entropy for the current frame is higher than the threshold defined for that block, then an abnormal event is indicated . Fig. 2 illustrates block-based processing framework using DCT in dynamic scenes.

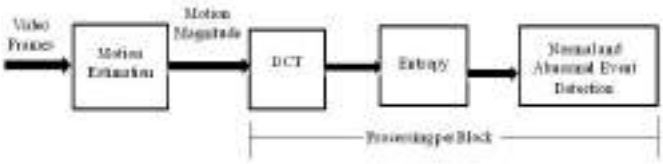


Fig. 2: DCT- Block based processing framework

IV. EXPERIMENTAL RESULTS

In our experiments, we use video dataset where several abnormal situations are simulated by a group of volunteers. These situations include running of several people to the middle and from the middle of the scene For an event to be considered normal or abnormal based on motion features, we compare the entropies for each block to the median averaged values over time to classify events into normal and abnormal using applying either DCT or DWT technique.

The video was divided into contiguous clips or frames. Anomalous video clips were separated from the rest of the video clips for testing. In the test data, anomalous clips were considered as positive examples and the non-anomalous clips were considered as negative examples.

In our implementation, we divide each frame into 4 blocks. For each block we calculate the entropy of the DCT or DWT coefficients of the motion vector magnitudes and then compute the median value over the first 100 frames. Based on experiments and evaluation, the threshold for the median entropy to classify an abnormal event is empirically set to 3 times the median value. Abnormality is detected for the whole frame is raised if abnormality is detected in any of the blocks.

Data Set 1: This video is of 25 sec, it is named as "Mov_0350" which is a footage recorded in college campus where we can observe that many people talking to each other and laughing, which actually indicates the normal event but all of a sudden some abnormal event took place and everyone started running in all the directions and this shows the abnormal event occurred in the video.

The video is divided into 95 frames. The video can be divided into any number of frames and it is depend on the user's choice. And the movement occurred in the video which indicates the abnormal is shown below and it is in the form of frames. Fig. 3a and 3b shows the frames of abnormal and normal situation in dataset1 video. Fig. 4a and 4b shows the motion estimation of abnormal and normal event in dataset1 video.

Fig. 3a : abnormal situation frame - DWT block based framework

Fig. 3b : Normal situation frame - DWT block based framework

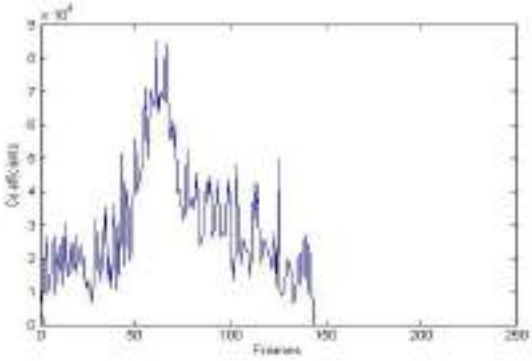


Fig. 4a: Graph indicates the motion estimation- Abnormal

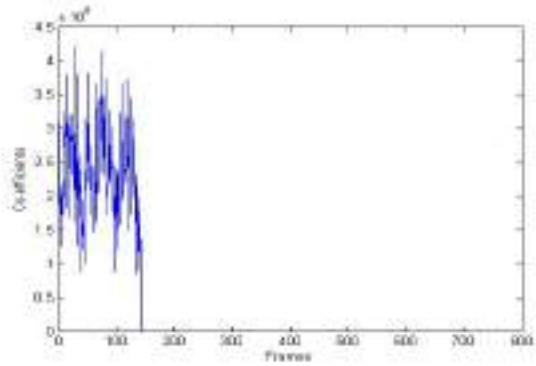


Fig. 4b: Graph indicates the motion estimation- Normal

Data Set 2: This video is of 1 min 40 sec, it is named as "Panic video" which is a footage recorded in train where we can observe that many people getting inside the train and settled down, which actually indicates the normal event but all of a sudden some event took place and everyone started running and get down from the train and after some time again everyone get inside the train and this shows the abnormal event occurred in the video.

The video is divided into 95 frames. The video can be divided into any number of frames and it is depend on the user's choice. And the movement occurred in the video which indicates the abnormal is shown below and it is in the form of frames. Fig. 5 shows the frames of abnormal situation in dataset 2 video. Fig. 6 shows the motion estimation of abnormal event in dataset2 video.

Fig. 5: abnormal situation frame - DCT block based framework

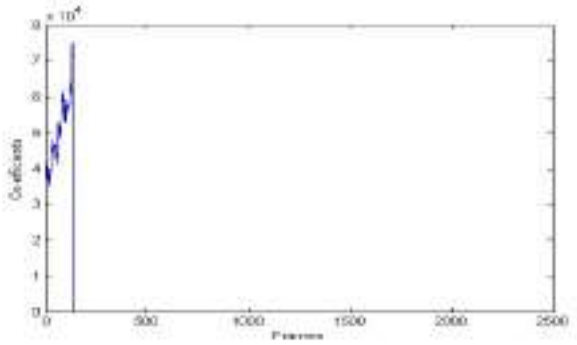


Fig. 6: Graph indicates the motion estimation

V. CONCLUSIONS

In this paper, we have developed a motion-context-based algorithm by applying DWT and DCT techniques to detect abnormal events in surveillance videos of a public place. We are implemented informative features based on motion and using threshold to detect abnormal events. In our approach, for classifying whether the current activity in the video is normal or abnormal, the entropy of the DWT or DCT-transformed motion magnitude is a reliable measure. Because the proposed methods are block-based, we can indicate exactly in which part of the frame the abnormal event takes place. Our proposed system detects the anomalies successfully and discriminate between Normal and Abnormal events in the given video data set.

REFERENCES

[1] L. Ovsenk, A. Kolesrov, J. Turn, Video Surveillance Systems", Acta- Electrotechnica et Informatica, Kosice: FEI TU, vol. 10, no. 4, pp.46-53, 2010.

[2] J. Li, Sh. Gong, T. Xiang, Global Behaviour Inference using Probabilistic Latent Semantic Analysis", the 19th British Machine Vision Conference (BMVC), pp. 193-202, Leeds, UK, September, 2008.

[3] Y. Benezeth, P.-M. Jodoin, V.Saligrama, C. Rosenberger, Abnormal Event Detection based on spatio-Temporal Co-occurence", in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2458- 2465, 2009.

[4] T. Chen, H. Haussecker, A. Bovyrin, et al., Computer vision workload analysis: case study of video surveillance systems", intel. Technology journal, vol. 9, no. 2, pp. 109-118, 2005.

[5] H.Boxton, Learning and understanding dynamic scene activity: A review", Image and Vision Computing, vol. 21, no. 1, pp. 125-136, 2003.

[6] N. Johnson and D. hogg, Learning the distribution of object trajectories for event recognition", Image and Vision Computing, vol. 14, no. 8, pp. 609-615, August, 1996.

[7] I. Junejo, O. javed and M. Shah, Multi feature path modeling for video surveillance", In Proc. of ICPR, no. 04, pp. 716-719, 2004.

[8] N. Ihaddadene, C. Djeraba, Real-time Crowd Motion Analysis, 19th International Conference on Pattern Recognition (ICPR), pp. 08-11, Tampa, Florida - USA, December, 2008.

[9] B. D Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision", In DARPA Image Understanding workshop, April, 1981.

[10] A., Wali, A.M, Alimi, Event Detection from Video Surveillance Data Based on Optical Flow Histogram and High-level Feature Extraction", 20th International Workshop on Database and Expert Systems Application, 2009.

[11] Syed Ali Khayam, (2003) The Discrete Cosine Transform : : Theory and Application.

[12] Gonzalez, R.C., R.E. Woods, S.L. Eddins, Digital Image Processing Using MATLAB, Gatesmark Publishing, 2nd edition, 2009.

[13] Chen T.s, Chang C. C.,and Hwang, K. S. F.,2002.Digital Image Processing, Taipei : Flag

[14] Weixin Li, Student Member, IEEE, Vijay Mahadevan, Member, IEEE, and Nuno Vasconcelos, Senior Member, IEEE, "Anomaly Detection And Localization In Crowded Scenes", January 2014.

Authors Biography

Mr. Puttegowda D, M.Tech (CSE) working as Associate Professor, Department of computer science and Engineering, ATME College of Engineering, Mysuru, Karnataka, currently pursuing research work on Video Mining in NAL research centre, Bangalore, affiliated to University of Mysore and has eleven years of teaching experience and two years of industry experience and published six international paper.

Dr. U N Sinha, the scientist, is very well known in the NAL, CSIR and the Indian scientific circles for his breadth and depth of knowledge in mathematics, fluid dynamics thermodynamics, parallel computing, atmospheric science and Sanskrit. Dr. U N Sinha obtained his engineering degree engineering in 1967 and his PhD from IIT Kanpur in 1976. In 1986 Flosolver project was born. It is creditable that Dr. Sinha and his team developed Flosolver Mk1, Indiaís first parallel computer and in 1993 Dr.U N Sinha and his team was first to complete the parallelization taking up of DSTís project to parallelize a global weather prediction model which was being used by NCMRWF for operational forecasts. Apart from his scientific achievements, Dr. Sinha has many other accomplishments to his credit: building a very good book collection in the library, teaching a very large number of students, taking good care of his colleagues.

Steganographic Data Embedding with Revocable Texture Synthesis

Akshay U.Joshi1, Sachin K S2, Vishal .M3, Aruna Kumara B4
1, 2, 3 UG Scholar, Dept. of CSE, RRCE, Bengaluru, 4Assistant Professor, Dept. of CSE, RRCE, Bengaluru
arun.bethur87@gmail.com

ABSTRACT — we propose a unique approach for steganography using a reversible texture synthesis. A texture synthesis method re-samples a smaller texture image that synthesizes a brand new texture image with the same native look and impulsive size. We weave the feel synthesis method into steganography to conceal secret messages. In distinction to exploitation associate existing cowl image to cover messages, our rule conceals the supply texture image and embeds secret messages through the method of texture synthesis. This permits us to extract secret messages and also the supply texture from a stego artificial texture. Our approach offers 3 distinct benefits. First, our theme offers the embedding capacity that's proportional to the scale of the stego texture image. Second, a steganalytic rule isn't doubtless to defeat our steganographic approach. Third, the reversible capability inherited from our theme provides practicality that permits recovery of the supply texture. Experimental results have verified that our planned rule will give varied numbers of embedding capacities, manufacture a visually plausible texture pictures, and recover the supply texture.

Index Terms—Data embedding technique, example-based method, revocable, steganography, texture synthesis.

I. INTRODUCTION

In the last decade several advances are created within the space of digital media, and far concern has arisen relating to steganography for digital media. Steganography a singular method of knowledge concealment techniques. It embeds messages into a bunch medium so as to hide secret messages therefore as not to arouse suspicion by a snooper. A typical steganographic application includes covert communications between 2 parties whose existence is unknown to a potential attacker and whose success depends on police work the existence of this communication. In general, the host medium employed in steganography includes meaningful digital media like digital image, text, audio, video, 3D model, etc. A large number of image steganographic algorithms are investigated with the increasing quality and use of digital images.

Most image steganographic algorithms adopt an existing image as a canopy medium. The expense of embedding secret messages into this cowl image is that the image distortion encountered within the stego image. This ends up in 2 drawbacks. First, since the dimensions of the quilt image is fastened, the additional secret messages that are embedded give additional image distortion. Consequently, a compromise should be reached between the embedding capability and also the image quality which ends within the limited capability provided in any specific cowl image. Recall that image steganalysis is associate approach accustomed discover secret messages hidden within the stego image. A stego image contains some distortion, and no matter however minute it is, this will interfere with the natural options of the quilt image. This leads to the second disadvantage as a result of its still potential that a picture steganalytic formula will defeat the image steganography and thus reveal that a hidden message is being sent in a very stego image.

In this paper, we tend to propose a unique approach for steganography using reversible texture synthesis. A texture synthesis method re-samples a small texture image drawn by a creative person or captured in a photograph so as to synthesize a brand new texture image with a similar native look and arbitrary size. We tend to weave the texture synthesis method into steganography concealing secret messages furthermore because the supply texture. Especially, in distinction to victimization an existing cowl image to cover messages, our algorithm conceals the supply texture image and embeds secret messages through the method of texture synthesis. This permits us to extract the key messages and also the supply texture from a stego artificial texture. To the most effective of our data, steganography taking advantage of the changeability has ever been conferred at intervals the literature of texture synthesis.

II. RELATED WORKS

Texture synthesis has received plenty of attention recently in computer vision and special effects. The foremost recent work has targeted on texture synthesis by example, within which a supply texture image is re-sampled using either pixel-based or patch-based algorithms to supply a replacement synthesized texture image with similar native look and arbitrary size.

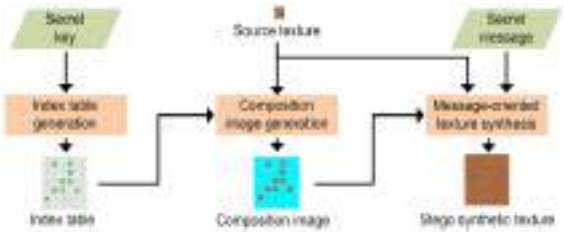
Pixel-based algorithms generate the synthesized image pixel by pixel and use spacial neighborhood comparisons to decide on the foremost similar pixel during a sample texture because the output pixel. Since every output pixel is set by the already synthesized pixels, any incorrectly synthesized pixels throughout the method influence the remainder of the result inflicting propagation of errors.

Patch-based algorithms take patches from a supply texture rather than a component to synthesize textures. This technique of Cohen et al. and Xu et al. improves the image quality of pixel-based artificial textures because the texture structures within the patches are maintained. However, since patches are affixed with alittle overlapped region throughout the artificial method, one must make an attempt to make sure that the patches accept as true with their neighbors.

III. SYSTEM ARCHITECTURE

We illustrate our planned technique during this section. First, we will outline some basic word to be utilized in our rule. The basic unit used for our steganographic texture synthesis is referred to as a "patch." A patch represents a picture block of a source texture wherever its size is user-specified. illustrates a diagram of a patch. We will denote the dimensions of a patch by its breadth (Pw) and height (Ph). A patch contains the central part and an outer part wherever the central part is observed as the kernel region with size of Kw×Kh, and therefore the half surrounding the kernel region is observed because the boundary region with the depth (Pd).

Next, we tend to describe the conception of the kernel block. Given a source texture with the dimensions of Sw×Sh we will subdivide the source texture into variety of non-overlapped kernel blocks, each of that has the dimensions of Kw×Kh, Let KB represent the gathering of all kernel blocks therefore generated, and ||KB|| represent the quantity of components during this set. We can employ the categorization for every supply patch kbi, i.e., KB= *KB*= {*kbi*| *i* =0 to ||*KB*||-1}. As an example, given a supply texture with the size of Sw×Sh =128×128, if we tend to set the dimensions Kw×Kh as 32×32, then we can generate ||KB||=16 kernel blocks.



The flowchart of the three-process message embedding procedure.

Each part in kb are often known {*kb*0, *kb*1…*kb*15}. We can expand a kernel block with the depth pd at all sides to produce a supply patch. The increasing method can overlap its neighbor block. indicates the boundary region of source patch sp4 once we expand the kernel block kb4 to overlap the kernel blocks kb0, kb1, kb5, kb8, and kb9. If a kernel block is found round the boundary of a supply texture, we operate the boundary mirroring using the kernel block's symmetric contents to supply the boundary region, as shown in for the kernel block kb4.

Similar to the kernel block, we are able to denote SP because the collection of all supply patches and SPn=||SP|| because the range of elements within the set SP. we are able to use the classification for every source patch spi, i.e., SP=spi to ||SP||-1}. Given a supply texture with the scale of Sw×Sh, we are able to derive the number of supply patches SPn using (1) if a kernel block has the size of Kw×Kh. In our paper, we have a tendency to assume the scale of the supply texture may be an issue of the scale of the kernel block to ease the complexity.

$$SP_n = \frac{S_w}{K_w} \times \frac{S_h}{K_h} \quad (1)$$

Processing Re-write Suggestions Done (Unique Article) Our steganographic texture synthesis formula should generate candidate patches once synthesizing artificial texture. The thought of a candidate patch is trivial: we've got a bent to use a window Pw×Ph then travel the provision texture (Sw×Sh) by shifting apixel on each occasion following the scan-line order. Let CP=i=0, 1, …, CPn-1 represent the set of the candidate patches where CPn=||CP|| denotes the quantity of elements in CP. We can derive CPn victimization (2).

$$CP_n = \|CP\| = (S_w - P_w + 1) \times (S_h - P_h + 1) \quad (2)$$

In our implementation, we use a flag mechanism. We first check whether or not the first supply texture has any duplicate candidate patches.

A. Message Embedding Procedure

In this section we will illustrate the message embedding technique. shows the three processes of our message embedding procedure. We will explain each process in the following sections.

1) Production of the Index Table

The first method is that the index table generation where we produce an index table to record the placement of the supply patch set SP within the artificial texture. The index table permits us to access the artificial texture and retrieve the supply texture completely. Such a reversible embedding style reveals one among the major advantages our planned algorithmic rule offers.

We 1st confirm the size of the index table (Tpw×Tph). Given the parameters Tw and Th , that square measure the breadth and the height of the artificial texture we have a tendency to shall synthesize, the number of entries during this index table is determined using (3) where total parenteral nutrition denotes the amount of patches within the stego synthetic texture. For simplicity, we selected applicable parameters for Tw, Th, Pw, Ph, and Pd, in order that the amount of entries is associate number. As associate example, if Tw×Th=488×488, Pw×Ph=48×48,

and Pd=8, then we will generate associate index table (12×12) containing 144 entries.

$$TP_n = T_{pw} \times T_{ph} = \left\lfloor \frac{(T_w - P_w)}{(P_w - P_d)} + 1 \right\rfloor \times \left\lfloor \frac{(T_h - P_h)}{(P_h - P_d)} + 1 \right\rfloor \quad (3)$$

When we distribute source texture to attain the style of reversibility, the source patches may be distributed in a rather sparse manner if the artificial texture contains a resolution that's much larger than that of the supply texture, On the contrary, the supply patches is also distributed in a rather dense manner if the artificial texture contains a resolution that is slightly larger than that of the supply texture. For the patch distribution, we avoid positioning a supply texture patch on the borders of the artificial texture. This will encourage the borders to be made by message-oriented texture synthesis, enhancing the image quality of the artificial texture. We additional outline the first-priority position L1 and also the second-priority position L2, for 2 kinds of priority locations where ||L1|| and ||L2||, dervied in (4), represent the quantity within the first-priority and second-priority positions, respectively.

Given the amount of patches SPn divided from the supply texture, the strategy of patch distribution is to distribute patches perfectly on the first-priority positions before posting patches on the second-priority positions. Supported the resolution of the synthetic texture, we'll have 2 cases: the thin distribution and dense distribution. These are represented below.

When the amount of supply patches is a smaller amount than or equal to the number of the first-priority positions (SPn≤||L1||), the patch will be distributed sparsely. within the security issue section we tend to describe some mathematical analyses of our algorithmic program. The analysis shows that the overall variety of patterns that the thin distribution offers is $C^{L1}_{SPn} x\ SP_n!$On the contrary, when the number of supply patches is bigger than the first-priority position (SPn>||L1||), the patch are going to be distributed densely. A mathematical analysis shows that the overall variety of patterns that the dense distribution offers is $C^{L2}_{SPn-L1} x\ SP_n!$

The index table has the initial values of -1 for every entry, which shows that the table is blank. Now, we'd like to re-assign values after we distribute the supply patch ID within the artificial texture. In our implementation, we tend to use a random seed for patch ID distribution that will increase the safety of our steganographic algorithmic program creating it harder for malicious attackers to extract the supply texture. As a result, the index table are going to be scattered with completely different values where we've got 9 supply patches (no. zero to 8) and one hundred thirty five blank locations with the initial price of "-1". During this index table, the entries with non-negative values indicate the corresponding source patch ID divided within the supply texture, whereas these entries with the worth of -1 represent that the patch positions will be synthesized by relating the key message within the message-oriented texture synthesis. Taking the higher than condition into thought, we are able to currently use the random seed Rs to disarrange the ID of the supply patches divided within the supply texture. As an example, if there ar 9 supply patches (SPn=9) and the artificial texture is synthesized with a complete variety of 144 patches (TPn=144), we are able to distribute the disarrayed 9 IDs of the supply patches leading to a thin distribution. Secret messages are going to be encoded within the remaining one hundred thirty five blank locations throughout the message-oriented texture synthesis.

2) Patch Composition Procedure

The second method of our formula is to stick the source patches into a worktable to supply a composition image. First, we establish a blank image as our bench where the scale of the worktable is adequate the artificial texture. By bearing on the supply patch IDs hold on within the index table, we then paste the source patches into the bench. Throughout the pasting method, if no overlapping of the supply patches is encountered, we paste the supply patches directly into the bench. However, if pasting locations cause the supply patches to overlap one another, we use the image quilting technique to reduce the visual object on the overlapped space.

3) Message-based Texture Synthesis Process

We have currently generated an index table and a composition image, and have affixed supply patches directly into the workbench. We'll insert our secret message via the message-oriented texture synthesis to provide the ultimate stego synthetic texture. The 3 elementary variations between our projected message-oriented texture synthesis and therefore the standard patch-based texture synthesis area unit delineated in Table I. The first difference is that the form of the overlapped space. During the conventional synthesis method, an L-shape overlapped space is normally accustomed confirm the similarity of each candidate patch. In distinction, the form of the overlapped space in our algorithm varies because we've affixed supply patches into the workbench. Consequently, our algorithmic rule must offer a lot of flexibility so as to address variety of variable shapes formed by the overlapped space.

The second distinction lies within the strategy of candidate selection. In standard texture synthesis, a threshold rank is usually given in order that the patch are often arbitrarily designated from candidate patches once their ranks area unit smaller than the given threshold. In distinction, our algorithmic rule selects "appropriate" patches by taking into thought secret messages. Finally, the output of the standard texture synthesis may be a pure synthetic texture

While the traditional texture synthesis algorithmic rule has an "L-shape" overlapped space, our algorithmic rule

might acquire another four shapes of the overlapped space. The texture space reveals a typical "L-shape" of an overlapped area, as shown in .However, once a close-by glued source patch has occupied the correct aspect of the operating location, this ends up in a "downward U-shape" of the overlapped space.. If a close-by glued supply patch has occupied the lower right corner of the operating location, this leads to a disjointed overlapped space containing AN "L-shape" and a tiny low however isolated half .Finally, if 2 near pasted supply patches have occupied the correct and bottom aspect of the operating location, this can contribute to AN "O-shape" of the overlapped space .

For each candidate patch inside the candidate list, one of the five shapes of overlapped space represented above can occur once referring to the synthesized space within the operating location. Thus, we can cipher the mean sq. error (MSE) of the overlapped region between the synthesized space and also the candidate patch. After all MSEs of the patches within the candidate list square measure determined. During this method, a section of the n-bit secret message has been concealed into the chosen patch to be affixed into the operating location.

B. Capacity Establishment

The embedding capability is one concern of the data embedding theme. Table II summarizes the equations we tend to described to investigate the embedding capability our algorithmic rule will offer. The embedding capability our algorithmic rule can give is related to the capability in bits that may be concealed at every patch (BPP, bit per patch), and to the quantity of embeddable patches within the stego artificial texture (EPn). Everypatch will conceal a minimum of one little bit of the key message; therefore, the edge of BPP can be 1, and also the the top capability in bits that may be concealed at every patch is that the bound of BPP, as denoted by BPPmax. In distinction, if we can choose any rank from the candidate list, the upper bound of BPP are $[\log_2(CPn)]$. The overall capability (TC) our algorithmic rule can give is shown in (5) that is that the multiplication of BPP and EPn. The quantity of the embeddable patches is that the distinction between the quantity of patches within the synthetic texture (TPn) and also the variety of supply patches subdivided within the supply texture (SPn).

$$TC = \mathrm{BPP} \times EP_n = \mathrm{BPP} \times (TP_n - SP_n)$$

Suppose we offer a supply texture Sw×Sh =128×128, and we will generate an artificial texture Tw×Th=488×488. We specify the patch size Pw×Ph=48×48 and therefore the boundary depth pd =8 pixels. this may cause the range of the BPP between one and 12. we will manufacture SPn=16 source patches and TPn=144 patches on the stego artificial texture. Thus, there square measure EPn =128 embeddable patches. If we take the BPP=12, then the full embedding capability is TC=1536 bit

C. Source/Supply Texture Recovery, Message Extraction and Message Authentication Procedures

The message extracting for the receiver side includes generating the index table, retrieving the supply texture, performing the texture synthesis, and extracting and authenticating the key message concealed within the stego synthetic texture. The extracting technique contains four steps.

Given the secret key command within the receiver side, an equivalent index table as the embedding procedure are often generated. The next step is that the supply texture recovery. Every kernel region with the size of Kw×Kh and its corresponding order with relevance the size of Sw×Sh supply texture are often retrieved by referring to the index table with the size Tpw×Tph. we are able to then organize kernel blocks supported their order, so retrieving the recovered source texture which can be precisely the same because the supply texture. Within the third step, we apply the composition image generation to stick the supply patches into a work table to produce a composition image by referring to the index table. This generates a composition image that's just like the one produced within the embedding procedure. The final step is that the message extraction and authentication step, that contains 3 sub-steps. The primary sub-step constructs a candidate list supported the overlapped space by referring to the present operating location. This sub-step is that the same because the embedding procedure, manufacturing an equivalent variety of candidate lists and their corresponding ranks.

The second sub-step is the match-authentication step. Given the current operating location Cur(WL) on the work table, we refer to the corresponding stego artificial texture at a similar working location Stg(WL) to see the stego kernel region SKw×SKh.

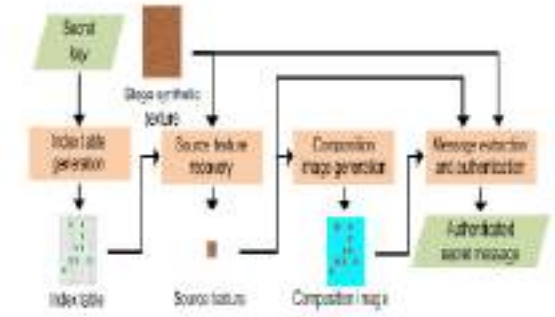


Fig. 7. The flowchart of the four-step message extracting procedure.

Then, based on this stego kernel region, we search the candidate list to see if there's a patch within the candidate list where its kernel region is the same as this stego kernel region. If this patch is out there, we see it as the matched patch, and denote it as MKw×MKh. Clearly, we will find the rank R of the matched patch, and this rank represents

the decimal worth of the key bits we sent within the stego patch when operational the feel synthesis within the message embedding procedure. However, if we cannot disclose any matched patch in the candidate list wherever the kernel region is the same because the stego kernel region, it implies that the stego kernel region has been tampered with, resulting in a failure of the message authentication. In this manner, we will authenticate and extract all of the key messages that are hid within the stego artificial texture patch by patch.

Our methodology is resistant against malicious attacks as long as the contents of the stego image don't seem to be modified. With some side information, as an example, our theme will survive the attacks of the image mirroring or image rotation by ninety, 180, or 270 degrees. However, if malicious attacks result in alteration of the contents of the stego texture image, the message authentication step can justify the genuineness of the key messages.

IV. CONCLUSIONS & FUTURE ENHANCEMENTS

This paper proposes a novel revocable steganographic algorithmic rule using texture synthesis. Given an artless supply texture, our scheme will manufacture an outsized stego artificial texture concealing secret messages. To the best of our information, we are the first ones that can fine weave the steganography into a conventional patch-based texture synthesis. Our technique is novel and provides reversibility to retrieve the first supply texture from the stego artificial textures, making doable a second round of texture synthesis if required. With the 2 techniques we have introduced, our algorithmic rule will manufacture visually plausible stego artificial textures even though the key messages consisting of bit "0" or "1" have an uneven appearance of possibilities. The given algorithmic rule is secure and strong against an RS steganalysis attack. We believe our proposed theme offers substantial advantages and provides an opportunity to increase steganographic applications. One potential future study is to expand our theme to support other kinds of texture synthesis approaches to enhance the image quality of the artificial textures. Another potential study would be to mix alternative steganography approaches to increase the embedding capacities.

REFERENCES

1. K. Thangadurai; G. Sudha DeviComputer "**An analysis of LSB based image steganography techniques",** Communication and Informatics (ICCCI), 2014 International Conference
2. P. Marwaha; P. Marwaha, "**Visual cryptographic steganography in images"**, Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference
3. M. F. Cohen, J. Shade, S. Hiller, and O. Deussen, "**Wang Tiles for image and texture generation**," ACM Trans. Graph., vol. 22, no. 3, pp. 287-294, 2003.
4. J. L. Rodgers and W. A. Nicewander, "**Thirteen ways to look at the correlation coefficient**," The American Statistician, vol. 42, no. 1, pp. 59-66, 1988.
5. D. Artz,"**Digital steganography: hiding data within data",** IEEE Internet Computing Year: 2001, Volume: 5, Issue: 3 Pages: 75 - 80, DOI: 10.1109/4236.935180
6. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "**Image quality assessment: from error visibility to structural similarity**," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, 2004.
7. D. Singla; M. Juneja "**An analysis of edge based image steganography techniques in spatial domain"**Recent Advances inEngineering and Computational Sciences (RAECS)", 2014

Gesture Recognition Based Car Gaming

Rubeena Muheeb , Dr. Parameshachari B D, Nagashree R N, Deekshith B N, Keerthikumar M, Rashmi P
Asst. Professor, Associate. Professor & HOD
Student , DoS in Computer Science,
Pooja Bhagavat Memorial Mahajana P.G. Centre, Mysuru
rubeena.muheeb@gmail.com, parameshbkit@gmail.com

ABSTRACT : This paper proposed a gesture recognition based car gaming proposed. This project developed based on the five skeleton gestures which will help to interact with the computer to play car game using kinect sensor. The five gestures namely Forward, Right, Left, Normal and Close. It will perform the operation with a slight delay. The kinect can be defined as a Natural Interaction (NI) device, which operates in the context of Natural User interface (NUI). NI refers to a concept where by the interaction with devices is based on human senses. Proposed method provides a suitable efficient and user friendly interface between human and computer for interaction with virtual game using human skeletal gestures.

Keywords - Gesture recognition, Natural Interaction (NI), Natural User Interface (NUI), Kinect Sensor, 3D Image

I. INTRODUCTION

Human motion analysis is gaining more and more attention in the field of human machine interaction. On one side, such popularity is caused by the fact that, existing devices become more sophisticated and combined with growing computational power has allowed to solve complicated problems. On the other side recently appeared number of more affordable devices which may be used as a part of relatively inexpensive systems. Human gesture recognition consists of identifying and interpreting automatically human gestures using a set of sensors. The documentation presents an up to-date review of the state-of-the-art in human gesture recognition which includes gesture representations, recognition techniques and applications. Several questions arise when tried to define the word "gesture".

The work Kinect is defined as a natural interaction (NI) device, which is used to operate with Natural user interface (NUI).Natural interface refers to interaction between human and the computer.

II. ABOUT KINECT SENSOR

The Kinect sensor consisting of several advanced sensing hardware**.** Kinect sensor mainly consisting of depth sensor a color camera, and a four-microphone array that provide 3D motion capture of human body, facial recognition, and voice recognition capabilities.

The arrangement of the Infrared (IR) projector, the color camera, and the IR camera is shown in the Figure1. The depth sensor mainly consisting of the IR projector combined with the IR camera, it is made up of monochrome complementary metal oxide semiconductor (CMOS) sensor. Kinect is defined as a motion sensing input device developed by Microsoft to interact with the Xbox *360* without the use of any a physical controller. It was originally designed as a natural user interface (NUI) for the Microsoft Xbox 360 video game console to create a new control-free experience for the user where there's no more need for an input controller. Kinect for Xbox 360 brings games and entertainment to life in extraordinary new ways with no controller required.

The IR camera measures the reflected light. Due to pattern recognition on the IR points and triangulation between the source and receiver, depth is measured. Prime Sense the company behind the technology of the Kinect talks of "Light Coding"-technology Prime Sense.

Figure 1 The infrared (IR) projector, IR camera, and RGB camera inside a Kinect sensor.

III. RELATED WORK

This paper consisting of recognition module and detection tracking. Gesture can originate from any bodily motion or state but commonly originates from the face or hand. Current focuses in the field include emotion recognition from face and hand gesture recognition. Many approaches have been made using camera and computer vision algorithms to interpret sign. Current focuses in the field include emotion recognition from face and hand Gesture recognition. Many approaches have been made using cameras and computer vision algorithm to interpret sign language. However, the identification and recognition of posture, gait, and human behavior is also the subject of gesture recognition technique.

The gesture based interaction interface being proposed here can be substantially applied towards many applications like Virtual Reality Sign Language and Games [1]. They have developed an interactive version of the game where gesture recognition based input technologies are successfully applied. They have implemented the game as playable and enjoyable using simple vision algorithms. The recognition system is composed of the server and client computers: the server recognizes the player's movements and the client the punch. Punch gestures detected by the client are fed into the server to control the game [2]. In this paper they proposed a novel method of using human body gestures depth image as gaming application input. Depth images have natural advantages over grayscale or color images in terms of robustness against illumination change, texture complexity, and background interference. The proposed method consists of three major components: depth image acquisition, mean shift established preprocessing, and HMM-based gesture recognition. They are validating their method by applying it to a boxing game scenario to distinguish boxing gestures such as dodge, jab, hook, and uppercut [3]. In this paper they explore the capacity of using skeleton information provided by Kinect for human posture recognition in a context of a heath monitoring framework. They conduct 7 different types of experiments with 4 types of features extracted from human skeleton. The obtained results show that this Kinect sensor can detect with high accuracy four interested postures (lying, sitting, standing, and bending) [4].In this paper, they propose a comparison of human gesture recognition using data mining classification methods in video streaming. In particular, they are interested in a specific stream of vector of twenty body-joint positions which are representative of the human body captured by Kinect camera. In the recognized gesture patterns the study are sit down, stand, and lie down. Classification methods chosen for comparison study are back propagation neural network support vector machine, naive Bayes and decision tree [5]. In this paper the system processes depth information to overcome the shortcomings of 2D vision system for the same application. It achieves left and right-palm tracking, and face and facial feature detection (eye, nose, and ears) detection.

The automation of the human body parts motion monitoring and its analysis in relation to the psychomotor exercise indicated to the patient and the storage of the result of the realization of a set of exercises free the rehabilitation experts of doing such demanding tasks [6].In this paper they proposes the topological representation encoded the intrinsic topology of the body's shape in a skeleton based structure, guarantying invariance to range, rotations and postural changes, and getting a high level of detail with a moderate computational cost. In the volumetric illustration, on the other hand, the postures were described in terms of 3D cylindrical histograms working within a wider range of distances in a faster way and also guarantying good invariance properties [7]. This survey paper proposes a real time implementation for a human skeleton recognition by Kinect that can be used for vision-based human interfaces. Using the low-cost device Kinect with its SDK tool kit gives us a possibility to resolve with ease some difficult problems encountered when working with conventional cameras. In particular, we are interested in a specific stream of vector extraction of twenty body-joint as a coordinate to identify skeletal structure of the human body captured by Kinect camera. The recognized gesture patterns from skeletal structure used to study various postures [8]. This paper proposes a real time implementation and novel methods for a hand-pose estimation that can be used for vision-based human interfaces [9].

IV. ARCHITECTURE DESIGN

The Kinect is an upcoming technology which basically looks like a webcam. It detects the 3D image representation of an object. It tracks the skeleton of the person standing in front of Kinect camera within a finite amount of distance.

The Kinect which operates in the context of Natural User Interface. Natural interface refers to a concept whereby the interaction with devices is based on human senses. The input skeleton gesture image is preprocessed in Kinect and the features of the image are extracted from the input. Data base is generated using sensor, gesture is recognized.

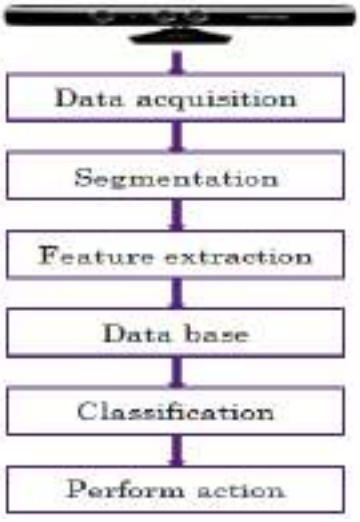


Figure 2 Game architecture

Segmentation is nothing but partitioning of the pixels from the image, in terms of homogenous and non-homogenous. It depends on the algorithm called Levenberg-Marquardt back propagation. Based on the segmentation feature extraction is done here the feature extraction is nothing but it is defined as extracting set of features or characteristics from input, and data base is generated. Based on data base, classification is done after that it will perform the action. The Figure 2 indicates the game architecture.

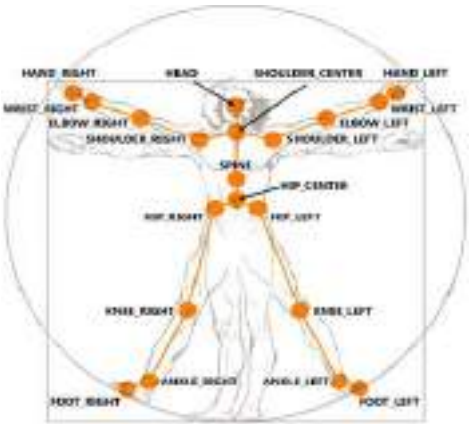


Figure 3 Human skeleton joints as reference points

The new version of Kinect with its SDK (Software Development Kit) containing the skeleton tracking tool. This tool provides us to collect 20 joint information about the skeleton. The joint information is collected in frames. For each frame, the positions of 20 points are estimated and collected. The 20 joints which are taken as a reference points is as shown in Fig 3. The first information is the index of the joints. Each joint has a unique index value. The second information is the positions of each joint in x, y, and z coordinates. These three coordinates are expressed in terms of meters. The x, y, and z axes are the body axes of the depth sensor.

From the skeleton tracked by the Kinect first it extracts the feature of joint positions. Since, each joint has 3 values and also 3 coordinates and the detected skeleton has of 20 joints. So, the feature vector has 60 dimensions.

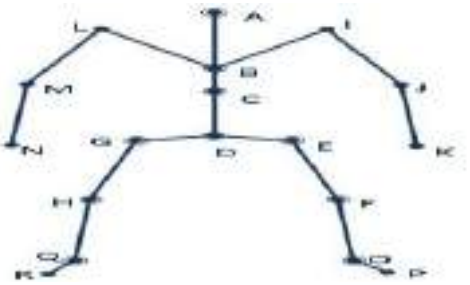


Figure 4 Important joints taken as reference points

The system can choose important joints for representing the postures. Other features can be derived based on joint position such as joint angles. Proposed design will analyze the recognition performance using these joints. Figure 4 shows the important point used as a reference points.

V. EXPERIMENTAL RESULTS

The image processing techniques used in the application for interactions with car game have been implemented in Matlab with the use of image acquisition toolbox in this image data acquisition functions are used.

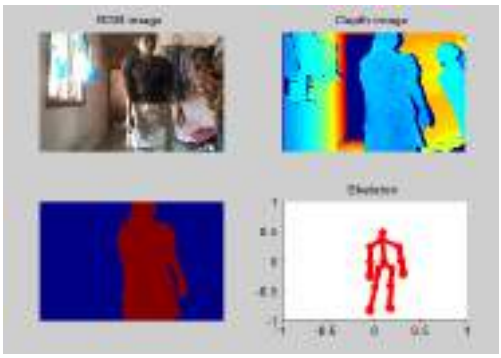


Figure 5 Forward gesture position

Figure 6 Right gesture position

In Figure 5 the proposed system identifies all the 20 coordinates with respect to pre-stored database this posture is identified as a Forward position. This can be achieved in the game by considering the condition that by considering the difference of z coordinate of shoulder centre and the z-coordinate of left hand and comparing this value with 0.5 Threshold and considering the difference of z coordinate shoulder centre and the x-coordinate of right hand and comparing this value with 0.5 Threshold.

Figure 6 the proposed system identifies all the 20 co-ordinates with respect to pre-stored database this posture is identified as a forward right position. By this gesture car will move in the forward and turning towards right direction. This can be achieved in the game by considering the condition that by considering the absolute value of difference of y - coordinate of left hand and the y-coordinate of right hand and comparing this value with 0.15 Threshold.

Figure 7 Left gesture position

In Figure 7 the proposed system identifies all the 20 co-ordinates with respect to pre-stored database this posture is identified as a Left position. By this gesture car will move in the forward and turning towards left direction. This can be achieved in the game by considering the condition by considering the absolute value of difference of y-coordinate of left hand and the y-coordinate of right hand and comparing this value with 0.15 Threshold.

VI. CONCLUSION AND FUTURE ENHANCEMENT

In the field of Human-Computer Interaction system design is majorly been in quality of interaction. Hence instead of designing regular interfaces, research branches have had focus on the concepts of intelligent and adaptive interfaces rather than command or action based ones. Developed algorithm is an interactive gesture for PC based car game called "Car Drifting" using a gesture recognition based interface.

The application of car game controlling uses four gesture through human skeletal gestures and verify the results with respect to different distances. The proposed algorithm operates properly from the range1000 to 3000 mm with a less delay. The proposed system shows different recognition rate with respect to distance and scenarios. This design provides a suitable efficient and user friendly interface between human and computer for interaction with virtual game using human skeletal gestures.

REFERENCES

1. Siddharth S. Rautaray and Anupam Agrawal, "Interaction with Virtual Game through Hand Gesture Recognition", 2011 International Conference on Multimedia, Signal Processing and Communication Technologies, 978-1-4577-1107-7/1 1©201 1 IEEE
2. Ji-Young Park, June-Ho Yi, "Gesture Recognition Based Interactive Boxing Game", International Journal of Information Technology Vol.12 No. 7 2006
3. Yong Wang, Tianli Yu*, Larry Shi*, Zhu Li*, "using human body gestures as inputs for gaming via depth analysis", 978-1-4244-2571-6/08/$25.00 ©2008 IEEE
4. Thi-Lan Le, Minh-Quoc Nguyen, Thi-Thanh-Mai, "Human posture recognition using human skeleton provided by Kinect", Nguyen International Research Institute MICA HUST - CNRS/UMI-2954 -Grenoble INP HANOI university of science and Technology VIET NAM, 978-1-4673-2088-7/13©2013 IEEE
5. OrasaPatsadu, ChakaridaNukoolkit and BunthitWatanapa, School of Information Technology King Mongkut's University of Technology Thonburi Bangkok, Thailand, "Human Gesture Recognition Using Kinect Camera" 978-1-4673-1921-8/12©2012 IEEE
6. D. González-Ortega~, F.J. Díaz-Pernas, M. Martínez-Zarzuela,M. Antón-Rodríguez, "A Kinect-based system for cognitive rehabilitation exercises monitoring", 0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
7. Giovanni Diraco~, Alessandro Leone, PietroSiciliano, "Human posture recognition with a time-of-flight 3D sensor for in-home applications", 0957-4174/$ - see front matter _ 2012 Elsevier Ltd. All rights reserved.
8. Manjunatha M B, Pradeep kumar B.P., Santhosh.S.Y, "Survey on Skeleton Gesture

Recognition Provided by Kinect" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 3, Issue 4, ISSN: 2278 – 8875, April 2014.

9. Jadhav Sheetal, Kale Payal, Jagtap Bhagyashree, Tamboli Azhar, "Hand Gestures Based Car Control System", International Journal of Advancement in Engineering Technology, Management & Applied Science. Vol. 2, Issue 1, Jan 2015.

ECG Signal Steganography Using Wavelet transforms

Asha N S, Anithadevi M D, Dr.Shivakumar K B, Dr. M Z Kurian
ashans.ec@gmail.com, anumdssit@gamil.com,shivakumarkb@ssit.edu.in, mzkurianvc@yahoo.com

ABSTRACT: Today, remote PoC (Point of Care) system is widely used to reduce to traffic in the hospitals. Privacy, security and confidentiality of the patient information are the major problem in the remote health care system. In is paper, we introduced encryption process on ECG steganography using wavelet transform. In proposed method, execute the time domain steganography for transmit the proficiency to conceal the patient information in any bit location with slightest error. And also frequency steganography is implemented in which five level wavelet packet decomposition is obtained. Data hidden inside the ECG signal through shared key and scrambling matrix. In this paper we have collected the ECG signal of patient from body sensors and physionet. We have hidden the physiological parameter into biomedical signal and produced watermarked biomedical signal. We have analyzed the energies of original and watermarked ECG signal using different wavelets such as coiflet, biorthogonal and symlets wavelet. From the result we concluded that energy of encrypted ECG signal using coiflet is higher than any other wavelet transform.

Key words: Confidentiality, ECG, energy, encryption, watermarked ECG signal.

1. Introduction

Now days, patients are suffering from various diseases especially related to the heart. With large traffic in the hospital patient can not able to get the appropriate treatment in time and the chance to they may get die. In order to solve this problem health care system uses remote PoC (Point of Care) system in the hospitals. Here, collect the patient physiological signal from the patient body through body sensors. Next, the collected signals are sent to the patient PDA (Personal Digital Assistant) for further diagnosis processing. Finally, the ECG signals and patient information as well as diagnosis report are sent to hospital server via internet. Doctors can check the biomedical signals and take a possible decision in case of an emergency. By using internet as the main communication tool that introduces security, privacy as well as data integrity. According to HIPAA (Health Insurance Portability accountability Act), the information sent through the internet should be protected and secured. In this paper, encryption based steganography technique is used .This method is not enough to secure the patient information. We have applied a new technique, which is obtained by wavelet transform.

Fig.1. shows the block diagram of ECG signal steganography scenario in Point of Care (PoC) system. In this method, ECG signals are collected from the different body sensor and physionet. Physiological parameter and patient personal information have taken from patient and sends to the patient PDA via Bluetooth. Steganography technique is implemented in the patient's PDA device.

Send this information to the hospital server via internet. All the doctors of the hospital can see the watermarked ECG signal only the authorized doctor can extract the secret information inside the host ECG signal.

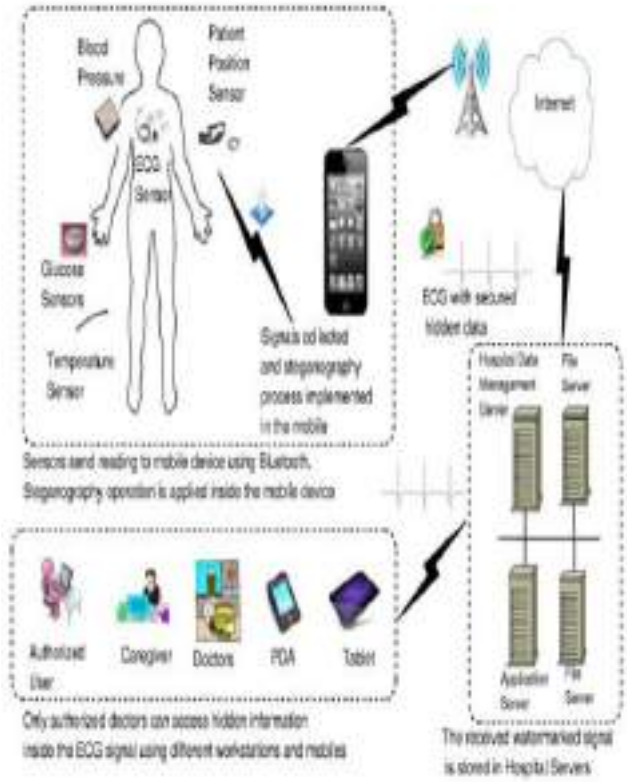


Fig.1. ECG steganography scenario in Point of Care (PoC) system where body sensors collect different readings as well as ECG signal and watermarking process implemented inside the patient's mobile device.

2. RELATED WORK

There are many techniques to secure the secret information. Navjot kaur and Usvir kaur proposed an audio watermarking using Arnold transformation with discrete wavelet transform (DWT) [3]. In this method, the secret information concealed inside the ECG signal by applying the DWT and DCT watermark scrambling with

Arnold transformation. After embedding extraction of secret information from ECG signal is completed. For checking the robustness, signal to noise ratio, mean square error and bit error rate is calculated. This technique was too lengthy and complexity level was high.

Nilanjan Dey, et al [4] proposed analysis of P-QRS-T components modified by blind watermarking method within the ECG signal for authentication in wireless telecardiology using DWT [4]. This method has two parts. In first parts, multi resolution wavelet transform based system is proposing for detection of P-QRS-T peaks complex from original ECG signal. P, Q, R, S, T peaks are detected and store over the whole signal. Time interval between two consecutive R-peak and rest peaks interval are measuring to check and detect abnormality of heart. To check the accuracy of P, Q, R, S, T components detection and interval measurement by processing and thresholding the original signal. The second part proposed the spread spectrum and discrete wavelet transform based watermarking. Watermarked signal is generating by DWT and compare distortion between watermarked ECG signal and original ECG signal but this method is not suitable for abnormal ECG QRS complex detection.

Golpira and Danyali [12] proposed a reversible blind watermarking for medical images based on wavelet histogram shifting. In this paper, medical images such as MRI are used as host signal. A 2-D wavelet transform is applied to the image. Then, the histogram of the high-frequency sub bands is determined. Next, two thresholds are selected, the first is in the beginning and the other is in the last portion of the histogram. For each threshold, a zero point is created by shifting the left histogram part of the first threshold to the left, and shifting the right histogram part of the second threshold to the right. The locations of the thresholds and zero points are used for inserting the binary watermark data. This algorithm performs well for MRI images but not for ECG host signals. Moreover, the capacity of these algorithms is low. Moreover, no encryption key is involved in its watermarking process.

3. PROPOSED METHOD

The sender side of the proposed steganography technique consists of four integrated stages as shown in fig.2. The proposed technique is designed to ensure information hiding with minimal distortion of the host signal. Moreover, this technique contains an authentication stage to prevent unauthorized users from extracting the hidden information.

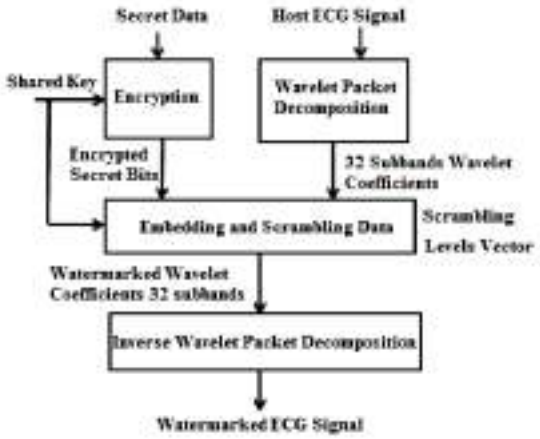


Fig.2. Blockdiagram of the sender steganography which includes encryption, wavelet decomposition, and secret data embedding.

3.1 Stage 1: Encryption

The aim of this stage is to encrypt the patient confidential information in such a way that prevents unauthorized persons – who does not have the shared key – from accessing patient confidential data. In RSA encryption method public key is used for encrypting the patient information and private is used for decrypting the patient information.

3.2 Stage 2: Wavelet Decomposition

Wavelet transform is the process of decompose the signal into coefficients representing frequency components of the signal at a given time. Wavelet transform can be defined as follows:

$$C(S, P) = \int_{-\infty}^{\infty} f(t)\psi(S,P)\,dt \qquad (1)$$

Where ψ represents the wavelet function. S and P are positive integer representing transform parameters. C represents the coefficients which is a function of scale and position parameters. In most of the application we use the discrete wavelet transform (DWT). DWT decompose the signal into two using band filters. The mathematical expression is defined as:

$$W(i,j)=\sum\sum X(i)\,\psi\, ij(n) \qquad (2)$$

Where W (i, j) represents the DWT coefficients, i, j shows the shift and scale parameters and ψ i,j shows wavelet basis time function. The result of the band filtering operation will be two signals, one is related to the high-frequency components and the other is related to the low-frequency components of the original signal. Low frequency component has important features of ECG signal and high frequency component contains noise. This process has implemented up to five levels.

3.3 Stage 3: Embedding Technique

Shared key and scrambling matrix is used in this stage. The patient physiological information is hide inside the host ECG signal by using scrambling matrix and private key. The Scrambling matrix can be defined as:

$$S = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,32} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,32} \\ \vdots & \vdots & \ddots & \vdots \\ s_{128,1} & s_{128,2} & \cdots & s_{128,32} \end{bmatrix}$$

Where, S represents 128×32 matrix and s represents number from 1 to 32. Two conditions are used for making matrix.

- Same row must contain different elements.
- Row of scrambling matrix must be different.

Embedding operation is shown in fig.4. At embedding stage, shared key is converted into ASCII codes. By the sequence fetcher row is read by the scrambling matrix.. After getting the first row, patient information is embedded inside the 32 sub band wavelet coefficients. In embedding operation, 32 sub bands wavelet coefficients are converted into binary form. Secret bits are replaced by LSB bits of wavelet coefficients and produced the stego ECG signal.

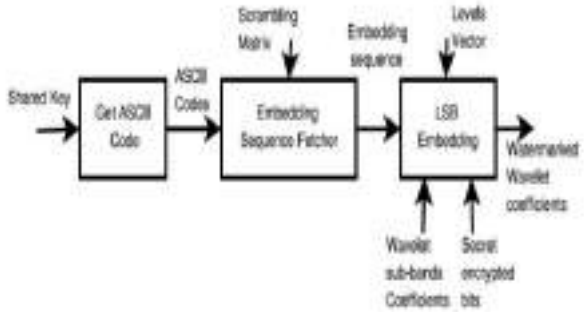


Fig.4. Block diagram showing the detailed construction of the watermark embedding operation.

3.4 Stage 4: Inverse Wavelet Decomposition

In this stage water marked wavelet coefficients are recomposed by applying inverse wavelet decomposition. After this stage, a new stego ECG signal is generated, which is similar to original ECG signal. It will be repeated until all the wavelet coefficients will be reconstructed. The inverse wavelet process will convert the signal to the time domain instead of combined time and frequency domain. Therefore, the newly reconstructed stego ECG signal will be very similar to the original ECG signal.

4. EXPERIMENTAL RESULTS

First we loaded the ECG signal and output of the original ECG signal is shown in the fig.5

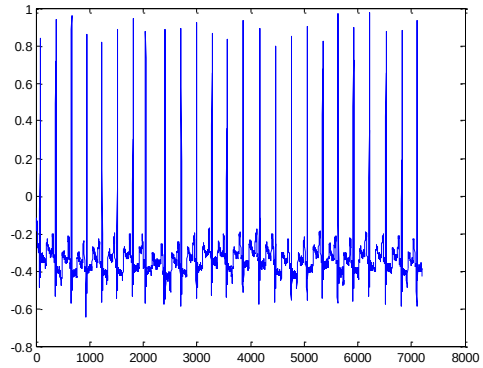


Fig.5. Host ECG signal

Afterwards, ECG signal has decomposed. Consequently, embedding process is implemented by using scrambling matrix and secret key. The bits of secret information are replaced with LSB bits of original signal. Subsequently, 32 sub bands watermarked wavelet coefficients are produced which is shown in fig.6.

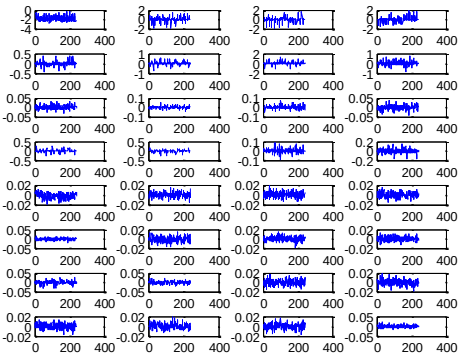


Fig.6. 32 sub band Wavelet packet coefficients

We have hidden the patient confidential information inside the ECG signal shown in fig.7.

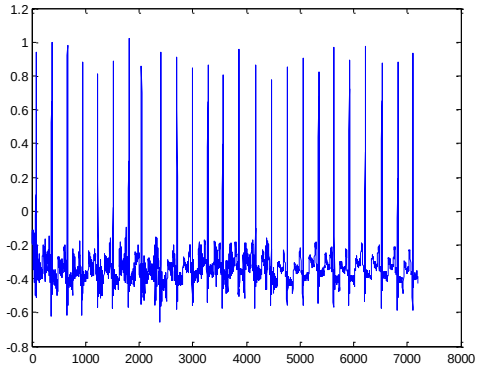


Fig.7. Data hide inside the ECG signal.

We have applied inverse wavelet packet decomposition to extract the patient information which is shown in fig. 8.

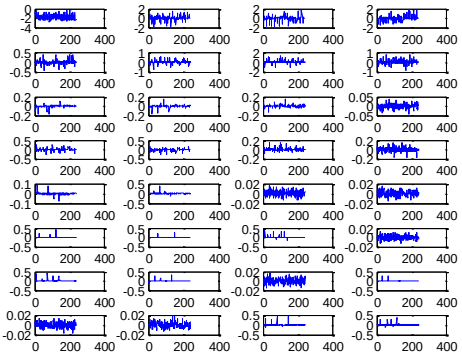


Fig.8. Inverse wavelet decomposition we get 32 sub band coefficients

Table 1 shows the energies of original ECG signal and watermarked ECG signal using different wavelets. E(original) re[resents energy of original ECG signal E(bior6.8) represents energy of encrypted ECG signal using Biorthogonal wavelet, E(sym4) represents energy of encrypted ECG signal using Symlets wavelet, E(coif5) represents energy of encrypted ECG signal using Coiflet wavelet.

Table 1

Energy of Original and watermarked ECG signal

E (original)	E (bior6.8)	E (coif5)	E (sym4)
559146 609	3.2632e+008	3.2632e+008	3.2631e+008
175010902	2.1168e+008	2.1168e+008	2.1165e+008
1.5903e+009	1.2073e+009	1.2073e+009	1.2072e+009
1.8880e+009	1.8693e+009	1.8695e+009	1.8695e+009

CONCLUSION

In the proposed method, sender steganography is implemented to encrypt the patient personal information and physiological parameters inside the ECG signal. A special range transform is implemented for shifting and scaling of the ECG signal which removes the negative value of ECG signal. Secret key is used for sending and receiving the message. Five level wavelet decomposition is applied to decompose the ECG signal. After decomposition, 32 wavelet coefficients are produced. Embedding operation is implemented by using scrambling matrix and shared key. In embedding process, secret bits of information are hidden inside LSB of cover signal. Finally 32 watermarked wavelet coefficients are produced. A new watermarked ECG signal is produced by inverse wavelet transform. The energies of original ECG signal and encrypted ECG signal are calculated by using different wavelets. We can observe that the energy of encrypted ECG signal using Coiflet Wavelet transform is higher than other wavelet transforms so Coiflet Wavelet transform can be used for encryption process in ECG steganography using wavelet transforms.

REFERENCES

[1] K. Malasri and L. Wang, "Addressing security in medical sensor networks," in *proc*. 1st *ACM SIGMOBILE Int workshop syst. Netw. Supp. Healthcare Assist. Living Environ*.,2007,p.12.

[2] H. wang, D. Peng, W. Wang, H. Sharif, H. chen, and A. Khoynezhad, "Resource- aware secure ECG healthcare monitoring through body sensor networks," *IEEE Wireless Commun., vol. 17,no. 1,pp.* 12-19, Feb. 2010.

[3] Navjot kaur and Usvir kaur, "An audio watermarking using Arnold transformation with discrete wavelet transform (DWT) and discrete cosine transform (DCT)."*et al/ International journal of computer science engineering* vol. 2,no 06, Nov 2013 .

[4] Nilanjan Dey, Sayantan Mukhopadhyay, Achintya, and Sheli Sinha Chaudhari, "Analysis of P-QRS-T componenets modified by blind watermarking technique within the ECG signal for authentication in wireless telecardiology using DWT". *International Journal of Image, Graphics, Signal Processing*, vol.4,no 7, July 2012.

[5] AymanIbaida and Ibarhim Khalil, "Wavelet based ECG steganography for protectingpatient confidential informationin point of care systems," IEEE Trans. Biomedical Engineering, vol. 60, no. 12, December 2013

[6] Dr PrenaMahajan and Abhishek Suchdeva, "A study of encryption algorithm AES, DES, AND RSA FOR security," *Global Journal of Computer Science and Technology Network, web and security* volume 13, issue 15 version 1.0 year 2013.

[7] A. Poularikas, *Transform and applications Handbook*. Boca Raton, FL, USA: CRC Press, 2006.

[8] Ganesh, G. Balasubhramanian, S.K, Jena, Pradhan," Simulation results for wavelet approximation," RGPA, no. 16, May 2012.
[9] Physiobank, physiotoolkit, and physionet: Components of a new research for complex physiological signals.
[10] Y. LIN. I. Jan, P. Ko, Y. Chen, J. Womg, and G. Jan, " A wireless PDA- based physiological monitoring system for patient transpoet". *IEEE Trans. Inf. Telchnol. Biomed*., vol. 8, no. pp. 439-447, Dec. 2004.
[11] F. Hu, M. Jiang, M. Wanger, and D. Dong, "Privacy-preserving tele-cardiology sensor networks: Toward a low-cost portable wireless hardware/software co design", *IEEE Trans. Inf. Technol. Biomed.*, vol. 1.1, no.6, pp. 619-627, Nov. 2007.
[12] A. Ibaida, I. Khalil, and F. Sufi, "Cardiac abnormalities detection from compressed ECG in wireless telemonitoring using principal components analysis (PCA)", in *Proc. 5th Int. conf. Intell. Sens. Netw. Inf. Process*.., Dec. 2010, pp. 207-212.
[13] W. Lee and C. Lee, "A cryptographic key management solution for HIPAA privacy/security regulations ", *IEEE Trans. Inf. Technol. Biomed*., vol. 12, no. 1, pp. 34-41, Jan. 2008.
[14] I. Maglogiannis, l. Kazatzopoulos, K. Delakouridis , and S. Hadjiefthymiades, "Enabling location privacy and medical data encryption in patient telemonitoring systems", *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 946-954, Nov. 2009.
[15] L. Marvel, C. Boncelet, and C. Retter, "Spread spectrum image steganography", *IEEE Trans. IMAGE. Process.*, vol.8, no.8 pp. 1075-1083, Aug 1999.
[16] A. De la Rosa Algarim, S. Demurjian, S. Berhe, and J. Pavlich- Mariscal, "A security framework for xml schemas and documents for healthcare", in *Proc. IEEE Int. Conf. Biomed. Workshop*, Oct. 2012, pp. 782-789.
[17] M. Li, S. Yu, Y. Zgeng, K. REN, AND W. Lou,"Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption", *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 131-143, Jan. 2013.
[18] S. Kaur, R. Singhal, O. Farooq, and B. ahuja, "Digital watermarking of ECG data for secure wireless communication ", in *Proc. Int. Conf. Recent Trends Inf. Telecommun. Comput.*, Mar. 2010, pp. 140-144.
[19] H. Golpira and H. Danyali, "Reversible blind watermarking for medical images based on wavelet histogram shifting", *in Proc. Int. Symp. Signal Process. Inf. Technol*., Dec. 2009, pp. 31-36.
[20] K. Zheng and X. Qian, "Reversible data hiding for electrocardiogram signal based on wavelet transforms", *in Proc. Int. Conf. Comput. Intell. Security*, Dec. 2008, vol. 1, pp. 295-299.
[21] A. Al-Fahoum, "Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure", *IEEE Trans. Inf. Technol. Biomed*., vol. 10, no. 1, pp. 182-191, Jan. 2006.

MANUAL SEGMENTATION OF HRCT LUNG IMAGE USING GAP FILLING TECHNIQUE

Dr.Punal M Arabi, Vamsha Deepa, Lakshmi Bai P
Department of Biomedical Engineering, ACSCE, Bangalore-74
arabi.punal@gmail.com, vamshi.deepa@gmail.com, lakshmibaip6@gmail.com

ABSTRACT -More than 3 million people died of COPD in 2012, which is equal to 6% of all deaths globally that year. Lung Disease diagnosis should be given war footing importance since Chronic obstructive pulmonary disease (COPD), a lung ailment contributed to 6% of all deaths globally in 2012. About 1.6 million people died in 2012 by lung cancer which is the most common cancer worldwide. Lung Disease diagnosis, treatment planning and monitoring the treatment efficiency all depend heavily on segmentation of lung tissues from the thoracic cavity. The techniques which use computer for lung disease diagnosis, treatment planning and monitoring the treatment efficiency depend on statistical analysis of lung tissue image data to make a decision. Hence segmenting the lung tissues from the thoracic cavity is the heart of the entire process. This paper presents a simple method for segmentation of lung with the help of hard threshold and gap filling technique. This method is experimented on a set of normal and ground glass lung images. The results obtained show that the proposed method is promising for segmenting lung from thoracic cavity.

Key words:- Gap filling technique, Hard thresholding ,HRCT (high resolution computer tomography) Lung Image ,Lung segmentation, Thoracic cavity.

I INTRODUCTION

Lung Disease diagnosis should be given paramount importance for Chronic obstructive pulmonary disease (COPD), a lung ailment claimed 6% of all deaths globally in 2012 [1] About 1.6 million people died in 2012 by lung cancer which is the most common cancer worldwide.[2] Segmentation plays a vital role in medical imaging since it is an efficient technique used to extract the region of interest from an image. Segmenting the lung tissue from the thoracic cavity is of particular interest for computer aided diagnosis and treatment monitoring of lung diseases. There several segmentation techniques proposed by researchers to segment lung tissues from thoracic cavity. Preeti Aggarwal et al [3] in their paper explored extensively the role of segmentation in medical image processing. Hu et al. [4] proposed a method which uses global thresholding technique to segment lung tissues. Automatic lung segmentation was suggested by technique Nihad Mesanovic et al.[5] for lung extraction; this is mainly a region growing technique. Adaptive region growing technique was the method used by Pohle and Toennies [6] for segmentation purpose in medical images. This paper presents a simple method for segmentation of lung with the help of hard threshold and gap filling technique. This method is experimented on a set of normal and ground glass lung images. The results obtained show that the proposed method is promising for segmenting lung from thoracic cavity.

II. METHODLOGY

In this paper, a new method is proposed for lung segmentation from CT (computer tomography) scan images. In the proposed method, CT images of normal lungs are acquired..The acquired image is converted to gray scale. So by converting the image to gray scale, the processing time is reduced and a faster algorithm is produced. The gray image is thresholded for 128. This is done in order the extract the lung from the thoracic cavity. Thus extracted lung image is obtained. Further the complete segmented lung image is obtained by gap filling technique.

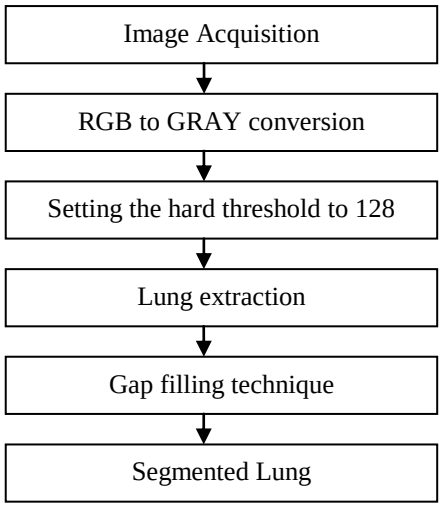


Fig 1:Flow diagram

III.RESULTS

For testing the proposed method ,a set of lung images of normal and lung with ground glass abnormality images are taken.Fig.2(a), 2(b), 2(c) are the images of normal lung, extracted lung tissues without gap filled, extracted lung tissues with gap filled. Similarly Fig.3(a), 3(b), 3(c) are the images of abnormal lung with ground glass structure, extracted lung tissues of it without gap filled, extracted lung tissues with gap filled. For Figures 2(a), 3(a) a hard threshold value of pixel intensity level 128 is applied. This resulted in Figures 2(b), 3(b), the extracted lungs without gaps filled. After applying gap filling technique on extracted lungs, the lungs with filled gaps are obtained and these tissues resemble closely the lung tissues in the original image.

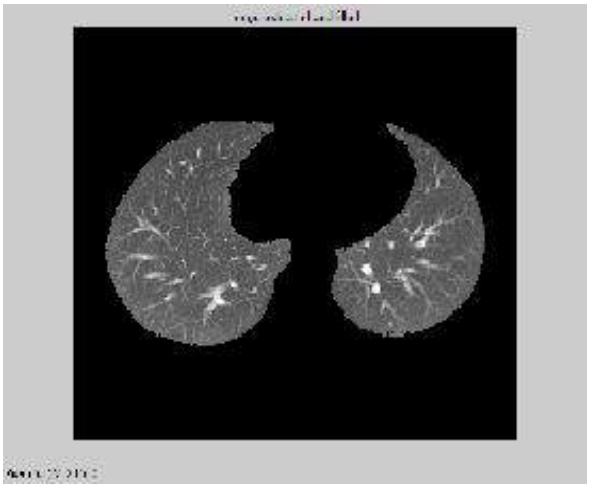


Fig2(c):Extracted lung with gap filled image

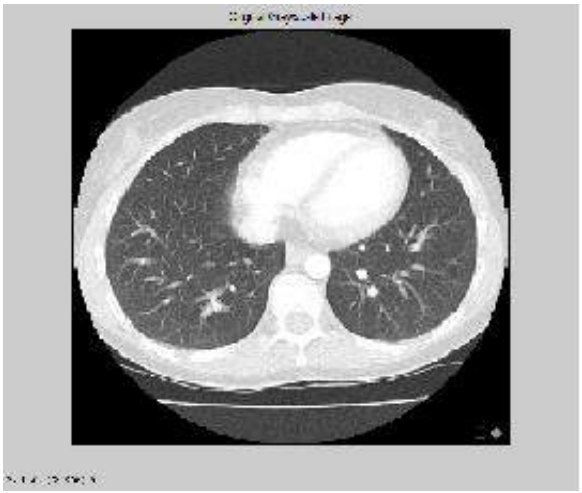


Fig2(a): Normal lung

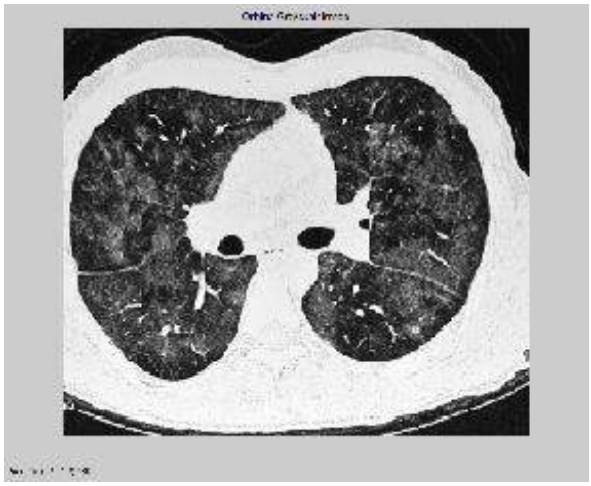


Fig3(a)Ground glass lung image

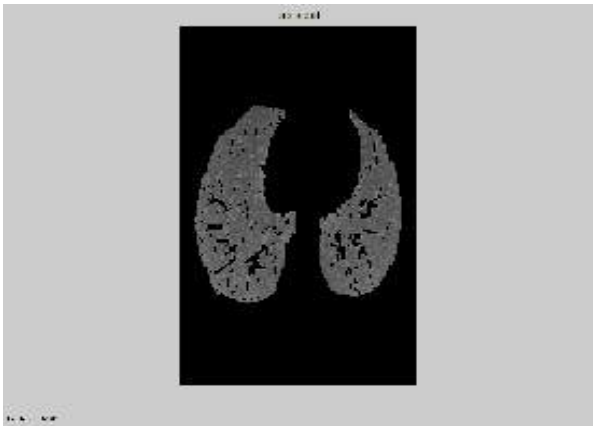


Fig2(b): Extracted lung without gap filled

Fig3(b): Extracted image without gap filled

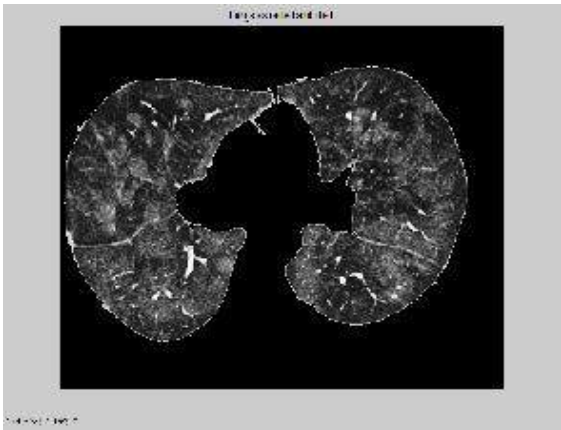


Fig3(c): Extracted lung with gap filled image

IV.DISSCUSION

Lung Disease diagnosis, treatment planning and monitoring the treatment efficiency all depend heavily on segmentation of lung tissues from the thoracic cavity. The techniques which use computer for lung disease diagnosis, treatment planning and monitoring the treatment efficiency depend on statistical analysis of lung tissue image data to make a decision. Hence segmenting the lung tissues from the thoracic cavity is the heart of the entire process. The method proposed in this paper is simple and depends on a hard threshold value of pixel intensity level 128.This level is chosen for the thoracic cavity of bright pixels and are of higher values than medium pixel intensity level 128. Hard thresholding pixel intensity level 128 removed the thoracic cavity along with all the other pixels having intensity value higher than 128.This leads to loss of information about lung blood vessels and pathological changes of lung tissues since blood vessels and pathological changes of lung tissues also have pixel intensity value higher than 128.This results in an error. The removed portions appear as gaps in the image. In order to fill up these gaps, gap filling technique is introduced which closes the gaps. The completely segmented lung image is now obtained.

V.CONCUSLION

Though the proposed hard thresholding lung segmentation method is tested on a set of lung images and yielded good results, it should be tested on many more images to know about the accuracy of the proposed method. In addition whether gap filling alters the pixel intensity information of the filled portion should be analysed.

ACKNOWLEDGEMENT

The authors thank the Management and Principal of ACS College of engineering, Mysore road, Bangalore for permitting and supporting us to carrying out this research work.

REFERENCES

[1] "Chronic obstructive pulmonary disease (COPD)"- Fact sheet N°315-January 2015
[2] www.lung.org
JOURNAL PAPERS:
[3] Preeti Aggarwal, Renu Vig, Sonali Bhadoria, C.G.Dethe, Sept 2011. "*Role of Segmentation in Medical Imaging:* A Comparative Study", International Journal of Computer Applications (0975 – 8887),Volume 29
[4] S. Hu, E. A. Hoffmann, J. M. Reinhardt, "*Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images*", IEEE Transactions on Medical Imaging, Vol. 20, pp. 490-498, 2001
[5] Nihad Mesanovic , Mislav Grgic , Haris Huseinagic , Matija Males , Emir Skejic , Muamer Smajlovic , "*Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm*"
[6] R.Pohle,K.D. Toennies, "*Segmentation of Medical Images Using Adaptive Region Growing*", Department of Simulation and Graphics, Otto-von-Guericke University Magdeburg, 2001.

AUTOMATED ATTENDANCE SYSTEM USING IMAGE PROCESSING

POOJA G.R, POORNIMA M, PALAKSHI S, M. BHANU PRAKASH VARMA, KRISHNA A N
Dept. of CSE, SJBIT, INDIA
krishna12742004@yahoo.co.in

ABSTRACT: Automated attendance system is a system which will record the status of the student; whether he/she is present in the class.In this paper we propose an automated attendance management system. This system, which is based on face detection and recognition algorithms, automatically detects the student using web cam when he enters the class room and marks the attendance by recognizing him. The system architecture and algorithms used in each stage are described in this paper. Different real time scenarios are considered to evaluate the performance of various face recognition systems. This paper also proposes the techniques to be used in order to handle the threats like spoofing. When compared to traditional attendance marking this system saves the time and also helps to monitor the students.

Index Terms—Face detection, Web cam, Automatic attendance, Performance.

I. INTRODUCTION

Time is gold as the famous line goes, time is very precious thing in today's fast changing, fast developing environment, through the different stages of civilizations man always tried to look for ways to make his work easier. Man was never satisfied, so the concept of automation came into existence. Automated attendance system is the advancement that has taken place in the field of automation replacing traditional marking attendance system. [1] The attendance monitoring system has made the lives of teachers easier by making attendance marking procedure a piece of cake. Face detection and recognition is an essential field in many applications, one which is Attendance Management System. This system includes detection of human face through a high definition camera where detection of images is done using a well-known algorithm called Viola Jones Algorithm [10].Feature Extraction is a method of capturing visual content of images for indexing & retrieval. Primitive or low level image features can be either general features, such as extraction of color, texture and shape or domain specific features. This paper presents an application of gray level co-occurrence matrix (GLCM) to extract statistical texture features for estimation of images. An automated attendance system which is developed using web cam consists of Image capture, face detection, database development, pre-processing, feature extraction, post-processing stages. This type of system can be ported in any academic environment.

II LITERATURE SURVEY

Normally in schools and colleges, the average student's count will be 50-80.Teacher has to mark student's presence for every hour. Traditionally, student's attendances are taken manually by using attendance sheet given by the faculty members in class, which is a time consuming event. Moreover, it is very difficult to verify one by one student in a large classroom environment with distributed branches whether the authenticated students are actually responding or not. Using conventional method of calling out names [5]takes approximately 5-10 minutes for marking attendance of entire class. It becomes complicated when strength is more. To overcome the traditional method, a finger print based attendance system came into existence.

Firstly employees fingerprint are scanned by software and an identity number is allotted as their enrollment .During the attendance time when employees impress their fingerprints against the scanner ,the system compares the new fingerprint patterns and connections between various points in the fingerprint with the enrollment database .A match is recorded as a knock exercising acquisition, processing, transmission, matching. Through this automatic system, time and manpower is reduced to great extent. In [6] a portable fingerprint device which can be passed among the students to place their finger on the sensor during the lecture time. The problem with this approach is that passing of the device may distract the attention of the students. A number of works related to Radio Frequency Identification [7] (RFID) based Attendance Systems exist in the literature. In RFID based system, students carry a RFID tag type ID card and they need to place that on the card reader to record their attendance. RS232 is used to connect the system to the computer and save the recorded attendance from the database. This system may give rise to the problem of fraudulent access. Iris [8] is the biometric that can be used for Attendance Systems. Iris Recognition: This recognition method uses the iris of the eye which is the colored area that surrounds the pupil. Iris patterns are thought unique. The iris patterns are obtained through a

video-based image acquisition system. Iris scanning devices have been used in personal authentication applications for several years. Systems based on iris recognition have substantially decreased in price and this trend is expected to continue. The technology works well in both verification and identification modes. This system uses iris recognition system that does capturing the image, iris recognition, extraction, storing and matching. The difficulty occurs to lay the transmission lines in the places where the topography is bad. To overcome all these problems, a real time face recognition system using web cam is proposed in which web cam is connected to computer, which will capture the faces of group of students. Using face detection technique, faces are extracted and processed using standard algorithms which is reliable, secure and fast.

III. METHODOLOGY

3.1 Viola Jones Framework Algorithm

This is a Paradigmatic method for Real time Face detection. Training is slow, but detection is very fast. The task of face detection in Viola Jones algorithm uses a 24x24 window as the base window size to start evaluating these features in any given image.

Fig.1 24 x 24 Base Window used by algorithm to crop face in given input image

Detection process has three key ideas

- The first is usage of **Haar Features** for the detection of features in the input image.
- The second is the introduction of a new image illustration called the **—Integral Image** which allows the features used by our detector to be computed very quickly.
- The third is an easy and efficient classifier which is built using the **AdaBoost learning algorithm** to select a small number of critical visual features from a very large set of potential features.
- The fourth contribution is a process for combining classifiers in a **—cascade** which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions.

Advantages:

- It is the most admired algorithms for face detection in real time.
- The main advantage of this approach is uncompetitive detection speed while relatively high detection accuracy, comparable to much slower algorithms.
- High accuracy, Viola Jones gives accurate face detection.
- Constructing a cascade of classifiers which totally reduces computation time while improving detection accuracy.
- The Viola and Jones technique for face detection is an especially successful method as it has a very low false positive rate.

Disadvantages:

- Extremely long training time.
- Limited head poses.
- Not detect black Faces.

3.1.1 *Haar features*

Haar features are similar to this convolution kernel which is used to detect the presence of that feature in the given image. Each feature results in a single value which is calculated by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle

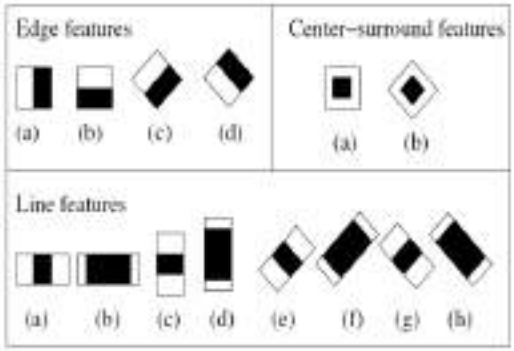


Fig.2 Different types of Haar Features

- If we consider all possible parameters of the Haar features like position, scale and type we end up calculating about 160000+ features in this window.


Fig.3 Line an e applied ose regions

Above images show use of different Haar features applied to eye regions, which cause redundancy in choosing Haar features.

3.1.2 *Adaboost*

It is used both to select the features and train the classifier. Weak learner is a single rectangle feature that best separates positive and negative examples; so weak classifier is a threshold single feature.

Initially, give equal weight to each training example. Iterative procedure to find the best weak learner for the current weighted training set and raise the weights of training examples misclassified by current weak learner. Compute final classifier as linear combination of all weak learners.

Advantages:

- AdaBoost is an algorithm which only needs two inputs: a training dataset and a set of features (classification functions). There is no need to have any a prior knowledge about face structure.
- Adaptive algorithm with relevant features under weak classifier, which is successively summed up to form strong classifier.
- Very simple to implement and program.

Disadvantages:

- The result depends on the data and weak classifiers.
- Quite slow training.
- Weak classifiers too complex leads to over fitting.
- Weak classifiers too weak can lead to low margins, and can also lead to over fitting.

Relevant features Relevant features Irrelevant features

Fig.4 Shows only selection of relevant features for weak classier

$$F(x) = \alpha_1 f_1(X) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + ...$$

3.1.3 *Cascading*

The basic principle of the Viola Jones face detection algorithm is to scan the detector many times through the same image, each time with a new size. Even if an image should contain one or more faces it is obvious that an excessive large amount of the evaluated sub-windows would still be negatives (non-faces). This realization leads to a different formulation of the problem:

Detection of eyes

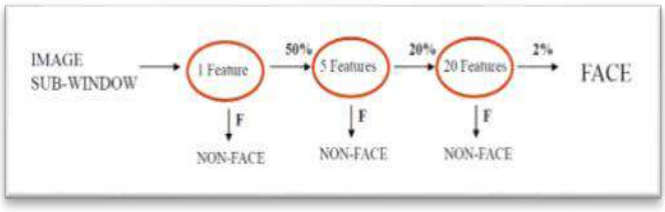


Fig.5 Instead of finding faces, the algorithm should discard non-faces.

The cascaded classifier is composed of stages each containing a strong classifier. The job of each stage is to determine whether a given sub-window is definitely not a face or maybe a face. When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a maybe-face is passed on to the next stage in the cascade. It follows that the more stages a given sub-window passes, the higher the chance the sub-window actually contains a face.

3.2 *Texture Analysis Based on Gray level co-occurrence Matrix (GLCM)*

Texture is a property that represents the surface and structure of an image or it can be defined as a regular repetition of an element or pattern on a surface. Textures of an image are complex visual patterns that are composed of entities or regions with sub-patterns with the characteristics of brightness, color, shape, size, etc.

Texture analysis characterizes the spatial variation of image pattern based on some mathematical procedures and models to extract information from it. One of the methods used for texture feature extraction was proposed by Haralick et al. known as **Gray-Level Co-occurrence Matrix (GLCM)**.

GLCM estimates image properties related to second-order statistics which considers the relationship among pixels or groups of pixels (usually two). A simple one-dimensional histogram may not be useful in characterizing texture features as it is a spatial property. Hence, this two-dimensional GLCM matrix is extensively used in texture analysis.

Gray-Level Co-occurrence Matrix (GLCM)

The GLCM, which is a square matrix, can reveal certain properties about the spatial distribution of the gray-levels in the texture image. It was defined by Haralick et al. in 1973. It shows how often a pixel value known as the reference pixel with the intensity value 'i' occur in a specific relationship to a pixel value known as the neighbor pixel with the intensity value j. So, each element (i, j) of the matrix is the number of occurrences of the pair of pixel with value 'i' and a pixel with value 'j' which are at a distance d relative to each other. The spatial relationship between two neighboring pixels can be

specified in many ways with different offsets and angles, the default one being between a pixel and its immediate neighbor to its right.

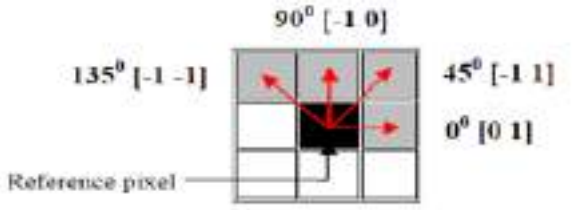


Fig.6 GLCM directions

There are 4 possible spatial relationships (0°; 45°; 90° and 135°).**Each element of the GLCM is the number of times that two pixels with gray tone i and j are neighborhood in distance d and direction θ.**

For 0° co-occurrence matrix, there are 2 occurrences of the pixel intensity value 1 and pixel intensity value 3 adjacent to each other (i.e. 1, 3) in the input image. Also, the occurrence of pixel intensity value 3 and pixel intensity value 1 adjacent to each other (i.e. 3, 1) is 2 times. Hence, these matrices are symmetric in nature and the co-occurring pairs obtained by choosing θ equal to 0° would be similar to those obtained by choosing θ equal to 180°. This concept extends to 45°, 90° and 135° as well with all these considerations, the GLCM matrix is calculated for each of the four possible angles and shown in fig.7.

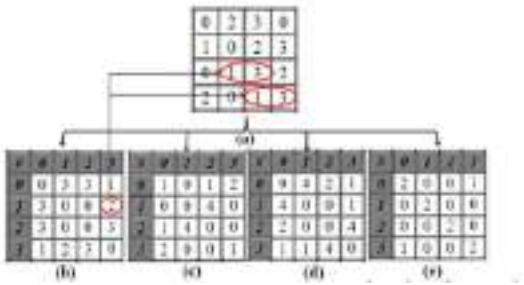


Fig.7 GLCM construction based on a (a) test image along four possible direction (b)0 (c) 45 (d) 90 and (e) 135 with a distance d=1. Here # represents the number of times.

IV RESULT AND DISCUSSION

4.1 *DATABASE TESTING*

The enrolment of the new student can be done in this stage. The images of the students are uploaded successfully.

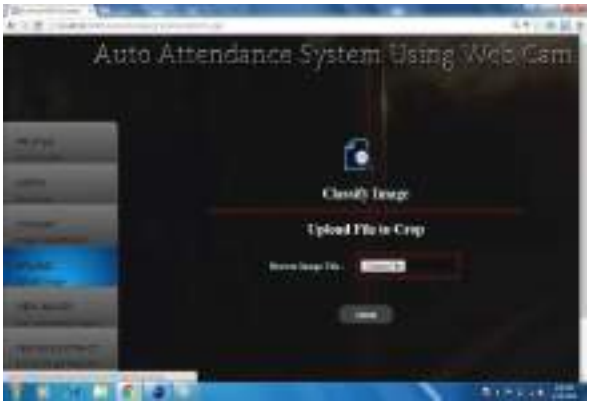


Fig.7 uploading of an image

Fig.8 Student list stored in database

4.2 *ATTENDANCE LIST*

The list of the students who attended the class are marked present

Fig.9 Student's attendance list

TABLE I: Performance analysis

Distance of object for correct recognition	5feet
Training time	670ms
Detection rate	90%
Recognition	80%

V CONCLUSION AND FUTURE WORK

An automatic attendance management system is a necessary tool for any learning management schools. Most of the existing systems are time consuming and require a semi manual work from the teacher or students. Our approach aims to solve the issues by integrating face recognition in the process. Even though this system has disadvantages like detecting large number of faces, there is much more room for improvement. Since we implement a modular approach we can improve different modules until we reach an acceptable detection and identification rate. The system can be enhanced in such a way that the accuracy, detection rate and recognition rate can be increased so that more number of students can be detected and recognized for those who are present in the class.

REFERENCES

[1] Shireesha Chintalapati, M.V. Raghunadh, "Automated Attendance Management System Based On Face Recognition Algorithms", Department of E and CE, *IEEE International Conference on Computational Intelligence and Computing Research 2013*.

[2] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of Human Faces," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 59, pp 748 – 760, May 1971

[3] Ming-Hsuan Yang, David J. Kriegman, NarendraAhuja, "Detecting Faces in Images: A Survey", 2002 IEEETransactions.

[4] Paul Viola, Michael J.Jones, "Robust Real-Time Face Detection" International Journal of Computer Vision Volume 57 Issue 2, pp. 137 – 154, May 2014.

[5] M. A. Fischler and R. A. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Transaction on Computer*, Vol. C-22, pp. 67-92, 1973.

[6] B. K. Mohamed and C. Raghu, "Fingerprint attendance system for Classroom needs," in *India Conference (INDICON),Annual IEEE*. 2012, pp. 433–438.

[7] T. Lim, S. Sim, and M. Mansor, "Rfid based attendance system," in*Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on*, vol. 2. IEEE, 2009, pp. 778–782.

[8] S. Kadry and K. Smaili, "A design and implementation of a wireless iris recognition attendance management system," *Information Technology and control*, vol. 36, no. 3, pp. 323–329, 2007.

[9] T. A. P. K. K. L. P. M. L. M. P. A. W. G. D. P. J. G..Roshan-Tharanga, S. M. S. C. Samarakoon, "Smart attendance using real time face recognition," 2013.

[10] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

Emotion Recognition: Detecting Emotions from Textual Documents, Blogs and Audio Files

Lakshmi. K, Harshitha.K.Rao[1], Revathi. V[2], Mrinal[3], Archana .T.P[4]
Asst. Professor, UG students[1234]
Dept of CSE,Vivekananda Institute of Technology, Bangalore-74
lakshmikumesh@gmail.com, harshiraok@gmail.com, revathi0402@gmail.com, mrinal.1401@gmail.com, archanabanakar94@gmail.com

ABSTRACT : With the rapid growth of the World Wide Web and due to increasing amount of textual information such as blogs, discussion forums, review sites available on the web. Now a days the area of emotion detection is considered as an active research field. Textual data has proven to be the main tool of communication in human-machine interaction and the use of social networking sites is one of the approaches for putting views of user. In order to make it as human and real as possible, this communication is constantly improving. In this paper, we propose a novel approach to recognize emotions based on keyword spotting technique along with the concept of ontology. An emotion estimation from the textual documents, data entered by user on blogs like twitter, along with the audio is taken and emotions is extracted from it. Because text, either written or spoken, is the unrivaled way to express one's feelings and emotions, it is the best quality to design for building efficient emotion recognition systems. However, researches on building text-based emotion recognition systems are progressing slowly compared to face or voice based systems. In order to improve the textual methods of communication such as text files, textual documents, tweets, it is needed to analyse the emotion of user by studying the input text and emotions of the user.

Keywords- Audio, Blogs, Emotions, Emotion Recognition, Keyword Spotting Technique, Ontology.

I. INTRODUCTION

Human emotion recognition by analyzing textual documents appear challenging but it is found many times useful and essential as well due to the fact that most of the times textual expressions are not only direct using emotion words but, it also includes words which does not express any emotions in text document.

Human Emotions may be expressed in many ways like person's speech, face expression and written text known as speech, facial and text based emotion respectively[1] .In case of recognizing emotion from a piece of text document or a blog, any human can do this better than a machine. Only problem is he or she takes time.

Recognizing user's emotions is a major challenge for humans and machine as well. On one hand, people may not be able to recognize or state their own emotions at certain times. On the other hand, machines need to have accurate ground truth for emotion modelling, and also require advanced machine learning algorithms for developing the emotion models. Hard sensing methods and soft sensing methods have been traditionally used to recognize the emotion of a user. With hard sensing methods, sensors provide the data sources that may be relevant to emotion recognition such as audio.

The detection of the emotion of the sentence is the most important step in emotion recognition. This process of extracting the text containing emotion deals with finding the emotion class from the sentence. As proposed previously, in order to find the emotion from nearly all sentences the method of extraction of emotional class is used. This detection of emotion class is divided into two approaches. First, detection of emotion class using affective word. Second, using context level processing of a sentence. The

emotion class is nothing but those words which distinctly denotes the emotional expressions of the sentence to a Particular emotion. Emotion words are taken as a feature. The features that represent the emotion of a sentence are derived by using the emotion dictionary. First the emotion words for each emotion class such as happy, sad, angry, fear are selected and then synonyms of these keywords are searched to prepare the affective words.

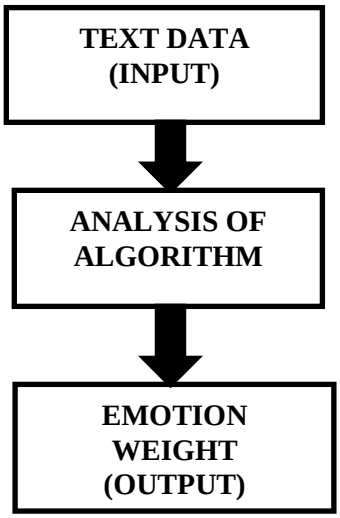


Fig-1 : Basic steps in Emotion Finder

Some of the methods that were used to find out the emotional state of a person are:

1.1 Keyword Spotting Technique

The keyword spotting technique can be described as the technique of finding occurrences of the keywords from a given text documents. Many algorithms to analyze a sentiment and emotions have been suggested in the past. In the context of emotion detection, this method is based on certain keywords that are predefined. These emotion words are categorized into keywords such as happy, angry, fear, disgusted, sad and surprise. Occurrences of these keywords can be found using some of the algorithms and based on that the given text document is assigned with an emotion class. "Today was a great day" is considered as positive, it is likely to assign the same classification For example a sentence like "Today wasn't a great day at all." Keyword spotting also depends on the actual affect of words and not for a sentence.

Sometimes, a sentence conveys more by their meaning instead of affect adjective words. For ex the following sentence "My husband just wants to take custody of my children away from me" states strong emotions, but uses no affect keywords, and therefore is ineffective.

1.2 Lexical Affinity Method

Detecting emotions based on related keywords is easy to use and straight forward method. Keyword spotting technique is extended into lexical affinity method which assigns a probabilistic 'affinity' for a particular emotion to arbitrary words apart from picking up emotional keyword.

For example, lexical affinity might assign a 70-percent probability to the word "accident" indicating a negative effect, as in "bike accident" or "hospitalized by accident. There are two main issues in this method. First, sentences such as "I escaped an accident" and sentences with other meanings such as "I met you by accident" creates a problem in lexical affinity, as they work only on the word level. Second, this method is often biased towards a particular type. Due to which it becomes difficult to build and reuse the same model for different systems.

1.3 Learning-based Detection

In machine learning methods, the emotion is detected by using classification approaches based on a training dataset. Based on previous history, similar kind of document which was used earlier is searched. Strapparava et al [3] developed a system that used several interpretations of Latent Semantic Analysis used to detect emotions in text when there is no affect word. However, this method has a low accuracy because it is not context sensitive and lacks the semantic analysis of the sentence. Burget R. et al. [4] proposed a framework that depends heavily on the pre-processing of the input data and assigning it using a classifier. The pre-processing was done at the word and sentence levels, and was used to calculate the relation between each term and each emotion class.

They achieved an average accuracy of 80% for 1000 Czech news headlines using SVM with 10 fold cross validation. However their method was not tested on English dataset. Also it is not context sensitive as it only considers emotional keywords as features. Dung et al exploited the idea that emotions are related to human mental states which are caused by some emotional events. This means that the human mind starts with initial mental state and moves to another state upon the occurrence of a certain event. They implemented this idea using Hidden Markov Model (HMM) where each sentence consists of multiple sub-ideas and each idea is considered an event that causes a transition to a certain state. By following the sequence of events in the sentence, the system determines the most probable emotion of the text. The system achieved an F-score of 35% when tested on the ISEAR dataset (International Survey on Emotion Antecedents and Reactions), where the best precision achieved was 47%. The low accuracy was mainly due to the fact that the system ignored the semantic and syntactic analysis of the sentence, which made it non-context sensitive.

1.4 Hybrid Methods

In hybrid methods, emotions are detected by using a combination keyword spotting method and learning patterns collected from training datasets, in addition to information from different sciences, like human psychology. Extracting emotions from a text that does not contain keywords belonging to any emotion is addressed as the problem by few works.

Wu et al. proposed an approach for emotion mining at sentence level based on detecting predefined semantic labels and attributes of the sentence, then emotions classification based on psychological patterns of human emotions called emotion generation rule (EGR).

II. LIMITATIONS

From above discussion there are few limitations [9]:

i) Ambiguity in Keyword Definitions:
Detecting emotion using keyword-based is a simple and easy way to detect associated emotion, but there could be multiple and vague meanings of a particular keyword, as the meaning of several keywords change according to their different usages. For e.g. "I met with an accident today" and "I escaped an accident today" have different emotions as the usage of "accident" is different in both the sentences.

ii) Incapability of Recognizing Sentences without Keywords:
Keyword-spotting technique always searches for some specific set of keywords. Therefore, sentences which does not contain any keyword shows that they do not belong to any emotion at all, which is obviously wrong e.g. "I cleared

the first round" and "Hooray! I cleared the first round" shows the same emotion (joy), but the sentiment in former sentence could not be detected without "hooray", if "hooray" is the only keyword to find the emotion.

iii) Lack of Linguistic Information:
Keyword-based approaches always search for some specific set of keywords. Therefore, lack of linguistic information does not help in finding any emotion. For e.g. "I passed my exam today" and "Hooray! I passed my exam today" imply the same emotion (joy), but in former sentence linguistic term like "hooray" is missing and hence recognizing emotion of this sentence would be difficult.

III. PROPOSED WORK

The proposed architecture is based on keyword based approach along with the concept of emotion word ontology. Ontology is an explicit specification of conceptualization. Emotion ontology is developed by an open source ontology editor called Protégé. Ontology contains almost all the words that contribute to a specific emotion. It is stored in a tree like structure. Here we use Emotion Word Ontology which consists of both words and emoticons. The emotion word ontology consists as follows:

- **Keywords:** Keywords include intensity modifiers and negations. Intensity modifiers are the words which specifies the intensity of an emotion such as very, too, exclamatory marks etc., Negations are the words which negate an emotion such as not, isn't etc.
- **Emoticons:** Emoticons are pictorial representation of a facial expression using numbers, letters, symbols and punctuation marks usually written to express a person's feelings. Even they are included in the emotion word ontology.
- **Data:** It consists of the dictionary words which expresses the emotion of a person.

Since here we use the negations and intensity modifiers the disadvantages of the previous traditional methods are overcome in this system.

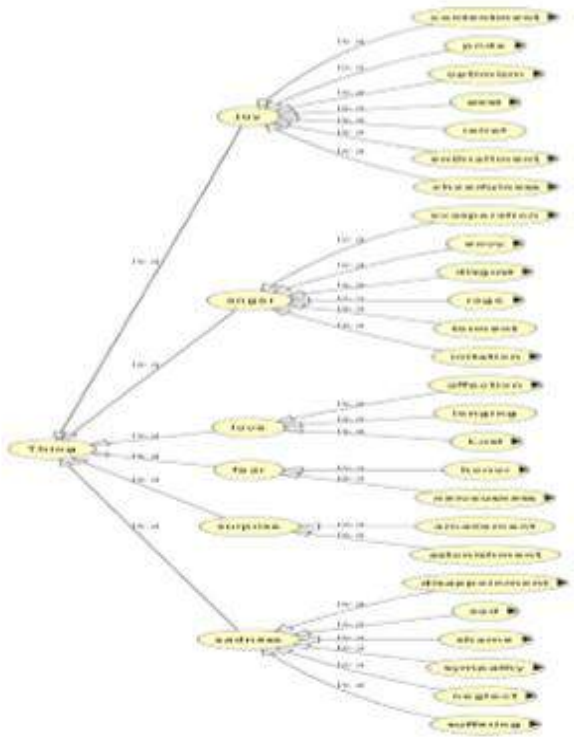


Fig-2 : Emotion Ontology

Algorithm:

Step 1: Read input text
Step 2: Split the input text into sentences
Step 3: For each sentence,
  3.1: Search for Punctuations at the end of sentence. If found add it to result.
Step 4: Split the sentence into words
Step 5: For each word
  5.1: If it is emoticon add it to results
  5.2: If it is an emoword.
    5.2.1: check the previous word for intensity modifiers and negations. And assign the emotion appropriately.
Step 6: Output will be emotion class.

Here we take text document as input and split it into sentences and then we check for punctuations such as "!", "?!" which denotes some emotions if found we add those emotions to the result and then we further divide the sentences to words and check for emoword (words that express some emotions) and add those to the result.

The algorithm makes use of three utilities:

- **Heuristic Utility:** It contains methods to find the coefficients of exclamatory mark, question mark, smileys etc.
- **Parsing Utility:** It breaks the paragraph into sentences and sentences into words.

- **Lexical Utility:** It checks whether the word is emoword or not.

Using the algorithm and the emotion word ontology we process three types of input data:

- **Text:** The text data is directly given as an input. The input for text can be taken from newspaper articles, documents, files etc.

- **Tweets:** The tweets are extracted from the twitter using a twitter handler and the tweets are given as input. We use the Alchemy API which is used to provide the web services by communicating with twitter to retrieve tweets.

- **Audio:** The audio input is converted to text using speech processing tool and then given as input which is then used for recognizing emotions.

IV. CONCLUSION

It has been seen that the Text-based emotion recognition plays an important role for humans to interact with computers while analysing the email for emotions, writing letters, emotion-based search engine that ranks documents according to the emotion requested by the user or giving feedback to any product in the era of new versions of web. Thus emotion detection from text acts as an important research field in affective computing. In this paper, existing system research of emotion recognition based on textual data is analyzed and drawbacks of existing methods are reviewed.

The emotion class (happy, sad, anger, fear, disgust, surprise) is recognized from the textual document apart from that recognizing the polarity of sentiment (positive/ negative/ neutral) from blogs takes place. Emotion recognition system architecture is proposed to improve detection of emotions in an efficient manner. Proposed system is based on keyword spotting technique that contains rich features of ontology. The advantage of our system is that it does not only detect the emotions from textual documents, but it also takes twitter blogs and audio as input. The audio input is first converted into the text using speech processing tool and then core algorithm is applied to find the emotions.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, "*Recognition of Emotional States in Natural human-computer interaction",* in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2009.

[2] Parrott, W.G, *"Emotions in Social Psychology",* in Psychology Press, Philadelphia 2001

[3] Carlo Strapparava, and Rada Mihalcea. *"Learning to identify emotions in text."* Proceedings of the 2008 ACM symposium on applied computing. ACM, 2008.

[4] Radim Burget, Jan Karasek, and Zdeněk Smekal. *"Recognition of emotions in Czech newspaper headlines."* Radioengineering 20.1 (2011): 39-47.

[5] C. Maaoui, A. Pruski, and F. Abdat, *"Emotion recognition for human machine communication",* Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, Sep. 2008

[6] Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo, *"Towards Text-based Emotion Detection: A Survey and Possible Improvements ",*in International Conference on Information Management and Engineering,2009.

[7] N. Fragopanagos, J.G. Taylor, *"Emotion recognition in human– computer interaction",* Department of Mathematics, King's College, Strand, London WC2 R2LS, UK Neural Networks 18 (2005) 389–405 march 2005

[8] C. Elliott, *"The affective reasoner: a process model of emotions in a multiagent system,"* in Doctoral thesis, Northwestern University, Evanston, IL, May 1992.

[9] D. S´anchez, M.J. Mart´ın-Bautista, I. Blanco, *"Text Knowledge Mining: An Alternative to Text Data Mining"* in 2008 IEEE International Conference on Data Mining Workshops.

[10] Z. Teng, F. Ren, and S. Kuroiwa, *"Recognition of Emotion with SVMs,"* in Lecture Notes of Artificial Intelligence 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 701-710

Performance Analysis of Different Filters with Various Noises In Preprocessing Of Images

Janaki K, Dr.M.Madheswaran
Research Scholar, AnnaUniversity, Chennai-25
karur.janaki@gmail.com, madheswaran.dr@gmail.com

ABSTRACT-- An image is considered as a collection of information and the occurrence of noises in the image causes degradation in the visual appearance of the images. So the information associated with an image tends to loss or damage. It should be important to renovate the image from noises for achieving maximum information from images. In this paper it has been seen that quality of image and the information in images affected by different types of noise. As a remedy, the visual quality and the information from the noisy image can be retrieved using different types of filters. In this work Gaussian noise, Speckle noise and random noise ,Salt and Pepper noise are being considered and it can be reduced using Gaussian filter, Average filter, Median filter and Adaptive Median filter. The experimental result shows the comparison and the various types of filters performance to de-noise the noisy images with mean square errors and PSNR values.
Keywords -Gaussian noise, Speckle noise,Salt and Pepper noise, Average filter, Adaptive Median filter.

I INTRODUCTION

Noise is an unwanted signal that interferes with the original signal and degrades the visual quality of digital image. The main sources of noise in the digital image are imperfect instruments, image acquisition process, transmission and compression, environmental conditions during image acquisition, insufficient light levels and sensor temperature, which affect the imaging sensor. Further noise in the image will be introduced when dust particles are present on the scanner screen. The image will be corrupt when there is interference of noise in the transmission channel. In the modern age, visual information is transmitted in the form of digital images by way of communication. But the image secured after transmission is often corrupted. Before it can be used in application, the received image requires processing to produce a visually high quality image. To eliminate noise from the original image, the digital image processing introduces image that have occurred at the time of image capturing due to the electronic and photometric sources. Degradation comes from blurring as well as noise. Identification of de-noising algorithm is application dependent. Therefore it is better to know about the type of noise present in the image to enable selection of the required de-noising algorithm. Still image de-noising poses a challenging problem for researchers as image de-noising causes blurring and introduces artifacts. Different types of image poses various types of noise. Different noise models are made use of to present different noise types. Hence de-noising method depends on the problem specific and relies on the type of image and noise model.Digital images find an important role in research and technology like geographical information systems in the field of medical science such as ultrasound imaging, X-ray imaging, computer tomography and MRI as well. Blurring is the type of bandwidth reduction of images due to imperfect image formation process such as relative motion between camera and original scene or by an optical system that is out of focus.

II LITERATURE REVIEW

1.To find different applications in image segmentation this paper [3] presented a new clustering-based segmentation. In this paper authors introduced a version of adaptive clustering based segmentation techniques. The basic concept behind this technique was to remove impulsive noises i.e., Salt-andPepper noise. Salt-and-Pepper noise contaminations, caused by errors in the image acquisition/recording/or transmissionare detectable at the minimum and maximum intensities. It is important to eliminate Salt-and-Pepper noise contained in theimage because its occurrence severely damage the information or data embedded in original image

2. In study of coherent imaging systems for e.g. ultrasound and laser imaging, multiplicative noise (also known as speckle noise)models are central. With respect to the standard Gaussian additive noise scenario these models have two additional layer of difficulties: 1) the noise is multiplied by (rather than added to) the original image; 2) the noise is not Gaussian, with Rayleigh and Gamma being commonly used densities [4].In this paper author performed a set of experiments and presented that their proposed method named MIDAL (multiplicative image denoising by augmented Lagrangian), yields state-of-the-art results in terms of speed as well as in denoising performance.

3.Author of this paper [5] proposed a recursive filter, called the Cluster-based Adaptive Fuzzy Switching (CAFSM)for removing impulse noise from digital images. This filter is composed of a cascaded easy to-implement impulse detector and a noise filter (for detail preserving restoration). During digital images acquisition process images are commonly get contaminated with impulse noise.

III PREPROCESSING FILTER WITH GRAYSCALE IMAGES

There may be a chance of noise to come in real time. Depends on real time scenario it is focused to add different kind of noise. Noise can be additive or multiplicative. Additive noise further classified into Gaussian noise, salt

and pepper noise, random noise. Multiplicative noise is speckle noise. Gray Scale Image or color image was taken for processing. If the input image is color, then it is converted into grayscale image. A color image has three components namely y-luminance, Cb-chrominance for blue, Cr- chrominance for red. After applying RGB to grayscale only y samples out of three input Y,Cb,Cr is obtained by skipping Cb,Crpixels.Noise was generated in the Sigma value ranging between 0.1 and 0.05.

Random number is generated between the range of 0 to 255. 0 represents value of black pixel, 127 represents value of gray pixel, 255 represents value of white pixel. It means that a random number will generate pixels within range of black to white pixels. Each pixel in the noisy image will have some random number between the range 0 to 255. N(X)x∑ is general noise function where ∑ indicates value ranging from 0.1 to 0.05 and n(x) is a value from 0 to 1. Highest value of sigma gives more noisy and lowest value of sigma gives less noise. Noise function is nothing but noise generation algorithm which will be varying for salt and pepper noise, speckle noise, random noise, Gaussian noise.

The generated noise was added to the original input image to obtain noisy image.0 to 255 is unsigned integer which is converted to double format which is ranging from 0 to 1 where 0 indicates black pixel value, 0.5 indicates gray pixel value, 1 indicates white pixel value. The entire image has to be converted into double format which haspixel values between 0 and 1. Since Noise is in double format, input image has to be converted to double format for summing of both noisy image and input image. Noise is added to input image to obtain noisy image.

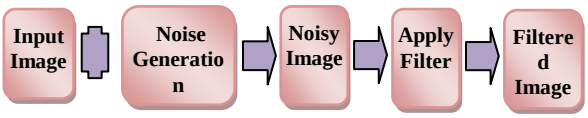


Figure 1. Block diagram of preprocessing filter for different noises

Filter is applied to noisy image to enhance certain features in the data and to remove noise from image, the image is subjected to different processing operations which involve changing the quality of an image in order to improve the pictorial information of an image for human interpretation, render the image should be more suitable for independent machine perception. The following equation is used to obtain noise free image from the noisy image.

Reconstructed image = ∑ h(m,n) × x(m,n)
where h(m,n) represents filter coefficients
and x(m,n) represents noisy image

Thus Filter is applied to the noisy image to reconstruct the noisy image. The filter coefficient values may be varying for different filters like Gaussian filter, average filter, median filter, adaptive median filter. This filtering process helps in maximize the clarity, sharpness of image and details of features of interest towards extraction of information & further analysis. The working of filter is depicted in the figure 1.

The original input image is compared with noisy image and Objective measurement parameters like PSNR-P1, MMSE-M1, SSIM-S1 and ENERGY RATIO-E1 were thus obtained. In the comparison of Original image with noisy image, the filter performance parameters have been PSNR,MMSE,ER,SSIM referred to as P1,M1,S1,E1 respectively

The noisy image is compared with filtered output image, Objective measurement parameters like PSNR-P2, MMSE-M2, SSIM-S2 and ENERGY RATIO-E2 were thus obtained. In the comparison of noisy image with filtered image, the filter performance parameters PSNR,MMSE,ER,SSIM are here shortly referred to as P2,M2,S2,E2 respectively.

The performance parameters of both comparison are and it is ensured that as P2>P1, M2<M1, S2>S1, E2>E1 which implied that the filtered output image gave good visual quality and less MMSE and appropriate SSIM and ER when compared to noisy image.

IV RESULTS AND DISCUSSION

Gaussian Noise (Amplifier Noise) noise has a probability density function pdf of the normal distribution. It is also known as Gaussian distribution. It is a major part of the read noise of an image sensor that is of the constant level of noise in the dark areas of the image.Gaussian noise is nothing but normal noise model. This model is additive in nature and follows Gaussian distribution. Each pixels in the noisy image is the sum of true pixel value in the noisy image and random, Gaussian distributed noise value. At each point, the noise is independent of intensity of pixel value. The PDF of Gaussian random variable is given by: $P(x) = 1/(\sigma\sqrt{2\pi})\ xe(x-\mu)2 / 2\sigma\ 2\ -\infty < 0 < \infty$ where: $P(x)$ is the Gaussian distribution noise in image; μ and σ is the mean and standard deviation respectively.

In many cases, noise in digital images is detected to be additive in nature with uniform power in the whole bandwidth and with Gaussian probability distribution. It can be referred to as Additive White Gaussian Noise (AWGN). Suppressing AWGN is rather difficult as it corrupts almost all pixels in an image. To eradicate noise without excessive smoothing of major details, a de-noising technique needs to be spatially adaptive. Depending on the noise model, various techniques are attempted. The wavelets naturally facilitates such spatially adaptive noise filtering due to the properties such as sparsity, edge detection and multi resolution. Median filters are good for Gaussian noise.

AWGN noise is added with original image of grayscale cameraman image as shown in Figure 3.2(left), and AWGN noisy image of grayscale cameraman image as shownin Figure 3.2(right)is generated this is done to compare original image and noisy image and thus to compute objective measurement parameters. These parameters have been found to be as PSNR=28.1191,MSE

= 101, ER=0.99438, SSIM=0.63213.

AWGN noisy grayscale cameraman image is reconstructed with Gaussian filter. AWGN noisy image as shown in figure 3.3(left) is compared with Gaussian filtered output image as shown in figure 3.3(right) to get objective measurement parameters and they have been found to be at the values PSNR=32.482,MSE = 68, ER=1.0098 ,SSIM=0.76095.

In this comparison, it is that the PSNR value is found to be increasing from 28 to 32.482 and MSE value getting reduced from 101 to 68, ER ranging between .99 to 1,0098 and SSIM getting increased from 0.63 to 0.76095 which is close to 1. Therefore preprocessing the input image with Gaussian filter gives better image quality compared to unprocessed image.

AWGN noisy imageis reconstructed with Average filter. AWGN noisy image as shown in figure 3.4(left) and Average filtered output image as shown in figure 3.4(right) are compared to find the values at which they change as exhibit fixed changes. To do so, objective measurement parameters are found to be PSNR=32.4262, MSE =210, ER=1.028, SSIM=0.78896.

In this comparison, it is noted that the PSNR value is increased from 28 to 32.4262 and but MSE is not reduced and is increased from 101 to 210 ER comes from .99 to 1.028 and SSIM is increased from 0.63 to 0.78896 which is close to 1. Also Average filter gives better objective measurements for image quality parameters than Gaussian filter. Therefore preprocessing the input image with Average filter gives better image quality compared to un-processed image. Since MSE not reduced with average filter, noisy image is preprocessed with another filter.

Using median filter, noise in AWGN noisy image is removed. In comparison with AWGN noisy image as shown in figure 3.5(left) with Median filtered output of grayscale cameraman image as shown in fiure 3.5(right), objective measurement parameters are computed as PSNR=35.5768, MSE =44, ER=1.0131 , SSIM=0.77695.

In this comparison, it is verified that the PSNR value is increased 28 to 35.5768 and MSE reduced from 101 to 44 ER comes from .99 to 1.0131 which is near to 1 and SSIM is increased from 0.63 to 0.77695 which is also close to 1. Also Median gives filter better objective measurements for image quality parameters than average filter. Therefore preprocessing the input image with median filter gives better image quality compare than average filter, Gaussian filter to some extent. Still little bit improvement needed for SSIM parameter. So input image is intended for preprocessing with another filter.

AWGN noisy image is reconstructed with Adaptive Median filter. In comparison of AWGN noisy image as shown in figure 3.6(left) with Adaptive Median filtered output of grayscale cameraman image as shown in figure 3.6(right), objective measurement parameters are computed as PSNR=36.497, MSE =144, ER=1.0131 , SSIM=0.78216.When comparing AWGN noisy image with Adaptive Median filtered outpu

Median filter give much improved PSNR value for and better image quality parameters than median filter. Therefore preprocessing the input image with adaptive median filter can be said to give better quality of image when compared to Gaussian,average,median filter. Thus for AWGN noise, performance of filters are analysed and it is ensured that AMF outperforms. For other noises also, performance of filters are analysed in the following sections.

Similarly speckle noise andrandom noise, salt and pepper noise added with original image of grayscale cameraman images and each and every noisy image is de-noised with four filtering techniques Gaussian filter, average filter, median filter and adaptive median filter. Thefilter parameters obtained for four kind of noisy images of grayscale image with various filtering techniques during preprocessing of image are shown in tables 1 to 4.

To estimate the performance analysis of Gaussian Filter, Average Filter, Median Filter, Adaptive Median Filter, the quality parameters such as PSNR, MSE, ENERGY RATIO, MAXERR, SSIM arecalculated for a given image. Performance of four different filters is tested against AWGN noise, speckle noise, random noise and salt n pepper noise which is generated or injected during image capturing or transferring image from scanner to computer.Experiments are conducted using Matlab 7.12.0.635(R20011a). Clinical dataset of grayscale image with image size of 256 x 256 pixels are taken as input image. Here 2D images are taken and dimension of original and noisy image are same. Noisy image is given as input to all the four different filters. The filtered outputs of each filter are given inFig.3 to 6.

Figure 2 Original Image Vs AWGNNoisy image

Figure 3 AWGN NoisyImage Vs Gaussian Filter output

Noise Type /Filters	PSNR	MSE	Energy Ratio	MAXERR	SSIM
Speckle Noisy Image	25.6057	178.8594	0.9928	43	0.0100
Gaussian Filter Output	33.1563	99.9928	1.0105	69	0.7024
Average Filter Output Image	32.8493	219.1137	1.0301	143	0.7457
Median Filter Output Image	29.8782	177.5321	1.0213	198	0.6795
AMF Filter Output Image	**34.7976**	**177.5321**	**1.0213**	**198**	**0.6901**

Figure 4 AWGN Noisy Image Vs Average Filter output

Figure 5 AWGN NoisyImage Vs Median Filter output

Figure 6 AWGN Noisy Image Vs Adaptive Median Filter output

Table 1Comparison of filter parameters of AWGN noise

Noise Type /Filters	PSNR	MSE	Energy Ratio	MAXERR	SSIM
AWGN Noisy Image					
Gaussian Filter Output Image	32.4820	67.4698	1.0098	65.3837	0.7610
Average Filter Output Image	32.4262	209.1787	1.0280	142.8303	0.7890
Median Filter Output Image	35.5768	143.5727	1.0131	192.9070	0.7770
AMF Output Image	**36.4970**	**143.5727**	**1.0131**	**192.9070**	**0.7822**

Table 2 Comparison of filter parameters for speckle noise

Table 3 Comparison of filter parameters for random noise

Noise Type /Filters	PSNR	MSE	Energy ratio	MAXERR	SSIM
Gray scale Image					
Random Noisy Image	25.0594	202.8352	0.9892	56.8186	0.5089
Gaussian Filter Output	29.7121	109.6790	1.0076	60.6594	0.6452
Average Filter Output Image	31.4936	221.0408	1.0276	138.0273	0.7342
Median Filter Output Image	33.0392	165.3227	1.0140	195.6642	0.7028
AMF Filter Output Image	**33.4870**	**165.3227**	**1.0140**	**195.6642**	**0.7083**

Table 4 Comparison filter parameters for Salt n Pepper noise

Noise Type /Filters	PSNR	MSE	Energy Ratio	MAXERR	SSIM
Salt N Peppers Noisy Image	25.31124	191.4062	0.991855167	247	0.010128
Gaussian Filter Output	40.93399	106.0367	1.009821224	172	0.827722
Average Filter Output Image	33.02128	221.3326	1.029521242	138	0.773086
Median Filter Output Image	44.31463	125.4986	1.011830071	199	0.87471
AMF Filter Output Image	**44.43203**	**125.4986**	**1.011830071**	**199**	**0.880131**

From the Tables 1 to 4, it can be understood that after the removal of noise from noisy image using different filtering techniques, the parameter values computed seen to varying. It is found that the filtered image by other filters having better PSNR values along with other parameters when

compare to noisy image. However, of all the filtering process, adaptive median filter is good and the output received is under to have a better quality of image because PSNR is high.

V CONCLUSION

In this paper four types of noises (Salt and Pepper, Gaussian, Speckle and random noise) have been added to input "Cameraman" image. It is observed that all noise causes degradation in the image quality which results in loss of information.In this analysis, Filters are used for picture quality evaluation. It should use objectivepicture quality measures such as PSNR,MSE, ER and SSIM.With respect to performance parameters PSNR, MSE, ER, MAXERR, SSIM, different four filters are analyzed for grayscale cameraman image and it is proved that Adaptive median filter outperforms as best de-noising technique in preprocessing of images .

FUTURE WORK

Filtering methods along with detection algorithms shows better results and once the filtering schemes are done in wavelet domain then efficient results will be found.

[1] http://en.wikipedia.org/wiki/Image_noise

[2] Rafael C. Gonzalez, Richard E. Woods, ―Digital Image Processing‖, 3rd edition, Pearson Education, 2008.

[3] SitiNorainiSulaiman, Nor Ashidi Mat Isa,―Denoising-based Clustering Algorithms for Segmentation of Low Level Salt-and-Pepper Noise-Corrupted Images,‖IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, pp. 2702-2710, November 2010.

[4] José M. Bioucas-Dias, Mário A. T. Figueiredo, ―Multiplicative Noise Removal Using Variable Splitting and Constrained Optimization,‖ IEEE Transactions on Image Processing, vol. 19, No. 7, pp. 1720-1730, July 2010.

[5] Kenny Kal Vin Toh, Nor Ashidi Mat Isa, ―Cluster-Based Adaptive Fuzzy Switching Median Filter for Universal Impulse Noise Reduction,‖ IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, pp. 2560- 2568, November 2010.

[6] Florian Luisier, Thierry Blu, Michael Unser, ―Image Denoising in Mixed Poisson–Gaussian Noise,‖ IEEE Transactions on Image Processing, Vol. 20, No. 3, pp. 696-708, March 2011.

[7] Kenny Kal Vin Toh, Haidi Ibrahim, Muhammad NasiruddinMahyuddin, ―Salt-and-Pepper Noise Detection and Reduction Using Fuzzy Switching Median Filter,‖ IEEE Transactions on Consumer Electronics, Vol. 54, No. 4, pp. 1956-1961, November 2008.

[8] PeihuaQiu and ParthaSarathi Mukherjee, ―Edge Structure Preserving 3D Image Denoising by Local Surface Approximation,‖ IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 8, pp. 1457- 1468, August 2012.

[9] M.Vijay, L.Saranya Devi, M.Shankaravadivu and M.Santhanamari "Image Denoising Based On Adaptive Spatial and Wavelet Thresholding Methods", IEEE-International Conference On Advances In Engineering, Science And Management, (ICAESM -2012) March 30, 31, 2012 161

[10] R.C. Gonzalez and R.E. Woods "Digital Image Processing", 2nd ed. Englewood Cliffs, NJ: Prentice-Hall; 2002.

[11] S.G.Chang, B.Yu and M.Vetterli, 2000, "Spatially adaptive wavelet thresholding with context modeling for image denoising," IEEE Trans. on Image Processing, Vol.9, No.9, pp. 1522-1531

[12] Kenny Kal Vin Toh,Nor, Ashidi Mat Isa," Noise Adaptive Fuzzy Switching Median Filter for Salt-and-Pepper Noise Reduction", IEEE Trans. Image Processing, Vol 17, No. 3, March 2010.

[13] Sheng-Fu Liang, Shih-Mao Lu, Jyh-Yeong Chang, and Chin-Teng (CT) Lin, "A Novel Two-Stage Impulse Noise Removal Technique Based on Neural Networks and Fuzzy Decision", IEEE Trans. Image Processing, Vol 16, No. 4, August 2008.

[14] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I.W. Selesnick, and C. S. Burrus, "Wavelet based speckle reduction with application to SAR based ATD/R," First Int'l Conf. on Image Processing, vol. 1, pp. 75-79, Nov. 1994.

[15] Robert D. Nowak, "Wavelet Based Rician Noise Removal", IEEE Transactions on Image Processing,vol. 8, no. 10, pp.1408, October 1999

[16] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filters under structural constraints," IEEE Trans. Signal Processing, vol. 43, pp. 591–604, Mar. 1995.

[17] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," IEEE Trans.Image Processing, vol. 3, pp.192–206, Mar. 1994.

[18] A. Ben Hamza, P. Luque, J. Martinez, and R. Roman, "Removing noise and preserving details with relaxed median filters," J. Math. Imag. Vision, vol. 11, no. 2, pp. 161–177, Oct. 1999.

Powerful Image compression technique on medical images

Janaki K, Dr.M.Madheswaran
Research Scholar, Research Supervisor, AnnaUniversity, Chennai-25
karur.janaki@gmail.com, madheswaran.dr@gmail.com

ABSTRACT-The digital image and its usage have increased may fold as the capability and utility of computers and networks have been increased rapidly. Against this background, there comes the need for improving image compression algorithms. Professional fields like medical imaging need acquiring huge amount of data every day and as a result of which compression reduces the file size. compressing data can save storage capacity, speed the file transfer and will decrease the costs for storage of hardware and network bandwidth.The need for an efficient image compression technique ever increasing because of the transmission & storage. For transmitting medical images over the internet,huge bandwidth is necessary.Number of images per diagnosis and resolution factor evenfor single patient makes the size of the images very large. So in compressing these medical images, there is an immense need for efficient compression techniques. For a given medical image noise is applied and noisy medical image is obtained. Noisy medical image is preprocessed with denoising technique then compressed. Compressed image is reconstructed then it is postprocessed and the result is compared with noisy image and picture quality parameters such as PSNR, MMSE, SSIM and ER, are obtained with extremely improved values. This paper deals scaling and hardthresholding techniques with HWT and highly acceptable compression ratio is achieved.

Keywords –Haar wavelet Transform, hard thresholding, scaling, adaptive median filter, entropy coding

I INTRODUCTION

An important area of biomedical engineering, especially for telemedicine are analysis and compression of medical images. Use of this will play a major role for diagnosis of sophisticated and complicated images through consultation. Transmission over networks at large distances for the use of telemedicine through medical image, it is necessary that they must be transmitted in compressed and secured form for the reliability.

Modern X-ray equipments (modalities) are introduced for use in the x-ray departments nowadays. These new scanners cannot exhibit films directly. On with the traditional x-ray equipments, the output results of mathematical calculations on signals detected during the scanning. The result from such scannings are digital images of the patients rather than analogue images. In the most traditional examinations the directly exposed films are not appearing.otherside,Plates of phosphor or selenium are put behind the patient, exposed with x-rays and subsequently put it in a reader along with the assistance of a laser beam converts the latent image toa digital signal. These digital images are forwarded to the above called laser cameras which print them out on film.

Data bases will have all these digital images instead of printing them on films. Actually this can be done but the number of PACS replacing fully the film based archives are few world wide and for several reasons. A medium size hospital with 50,000 X-ray examinations per annum produce equivalent of Three to four gigabyes per day. The health authorities may normally want some years production is recorded. For example in Denmark, national board of health stipulates that 5 years production has to be stored. It is very difficult to hookup digital X-ray equipment into network for image distribution and archival due to the lack of standards. As majority of the few completely filmless departments in the world is set up with equipment from only one vendor, X-ray devices, network, archive, everything result a very inflexible and also very expensive solution.

Remedial of the situation has come from major advances during the last few years. Readily availability of ATM networks starting at 155 megabits per second and also as well as network adapter for workstations, PC's and MAC's are found now. Defining of an internationally approved, complete protocol for medical image communication DICOM exists. Possibility for choosing the different component from different vendors based on a department completely on this standard. Archives capacity is still a problem. It is very costly to achieve decent performance. Solution will be in image compression even this problem is solved. Several

algorithms like predictive coding, sub-band coding, vector quantization, segmented image coding, neural networks and fractal coding have been proposed in the literature until now.The wavelet transform has got significance due to its manifold characters like high compression ratio, lossless to lossy, multiresolution in nature and applicable to various medical images.
Any noise like 'Gaussian noise' or 'salt and pepper noise' or 'moisture noise' or 'speckle noise' is added with a given medical image and noisy medical image is generated. Adaptive median filter is helping in the preprocessing of noisy medical image. Incompressed part, preprocessed medical image is partitioned into 8x8 blocks of pixels. For each and every block, gamma scaling with HWT and entropy coding is applied to obtain compressed image. The objective parameter CR is computed for compressed image. In compared with input noisy medical image and picture quality parameters like PSNR, MMSE, SSIM, ER, are taken along with extremely improved values.

II RELATED WORK

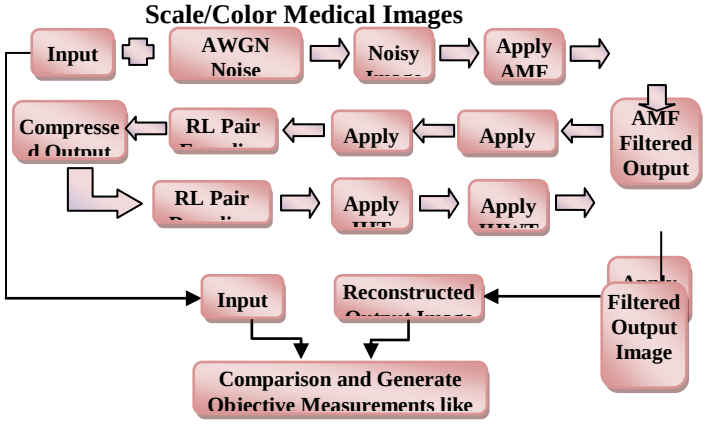
Sukhwinder Singh, Vinod Kumar, H.K. Vermahas Jointly proposed a novel technique for medical image compression called adaptive threshold-based block classification. As the result of this, CT, an X-ray and ultrasound image are used to evaluate the performance and compares the derived results to the JPEG respective to the quality indices. [1]

J. Jyotheswar, SudiptaMahapatra has proposed a paper on efficient FPGA implementation of DWT (Discrete Wavelet Transform) and modified SPIHT (set partitioning in hierarchical trees) for lossless image compression. As the result shows that, the algorithm promotes good compression ratio and better peak-signal-to-noise ratio (PSNR) with 3D medical images. [2]

A new technique of 3D wavelet transform was proposed by Gregorio Bernabe, Jose M. Garcia and Jose Gonzalez for medical videos . Result shows, a good compression ratio and a good quality of the reconstructed video. Also when comparing this scheme with MPEG-2 and EZW, promotes better compression ratios of 119% and 46% respectively for the same PSNR

decompression part, run length decoding followed by inverse thresholding along with upscaling, IHWT is applied on compressed image. Finally adaptive median filter as post-processing filter is applied on decompressed medical image to improve picture quality. Post-processed image of Decompressedmedical image is the peak signal-to-noise ratio (PSNR) value. [4]

III Image Compression-Decompression For Gray Scale/Color Medical Images



Yen-Yu Chen has designed a novel medical imagecompression technique called DCT(Discrete CosineTransform) based subband decomposition and modified SPIHT (set partitioning in hierarchical trees) data organization. Results showed that the quality of the reconstructed medical image has been

Figure 1Block Diagram of Image Compression-Decompression

An Gray Scale is taken Salt n Peppers Noise generator generates with range of Sigma value from 0.1 to 0.05.generated noise is added to the original brain image and salt n pepper noisy image is obtained as shown in figure 6.2.Adaptive Median filter is applied to the noisy image and filtered brain image is obtained as shown in figure 3.A preprocessed AMF filtered output image is taken and partitioned into 8x8 pixels.For every block of 8x8 matrices, HWT is applied and coefficient matrix is obtained.scaling techniques with high factor – Ngamma is applied on HWT output matrix.Hardthresholding is applied to scaled HWT output matrix.Entropy Coding is applied to scaled,thresholded HWT output and compressed output is obtained.

Run length decoding is applied on compressed output.Non-zero elements of the compressed stream are obtained and the better compression ratio is evaluated with reference to original stream. Inverse Hard thresholding followed by upscaling is applied on decoded output. Inverse HWT is applied on Inverse hardthresholded , upscaled, decoded output to reconstruct the compressed image.Reconstructed image is postprocessed with Adaptive median filter.Decompressed , postprocessed image is compared with noisy image to compute objective measurements parameters.Steps 1 to 15 are depicted iblock diagram as shown in figure 1. It is verified that Ngamma high factor provides best possible reduced number of digits for gray scale.Conventional Run Length pair decoding is applied to the compressed stream.Inverse Hard threshold is apply then Inverse HWT is applied followed by AMF and obtained reconstructed output image.It's ensured that this technique is providing better compression ratio and better image quality.It also ensured for all other noises like AWGN, Random Noise and Speckle nois**e.** It also applicable for different image file formats like jpeg, bmp, tiff, etc **.**

IV RESULTS AND DISCUSSION

Salt n pepper is added with original grayscale brain image in figure 2(left), and Salt n pepper noisy grayscale brain image is generated as shown in figure 2(right) in which original grayscale brain image is compared with salt and pepper noisy grayscale brain image and objective measurement parameters are computed as PSNR=24.1446,MSE = 251,ER=0.993472 , SSIM=0.048611

Figure 2 original Image Vs Salt n Peppers Noisy Image

In comparison of Original grayscale brain image with Salt n Peppers grayscale brain image Noisy Image, poor picture quality parameters are obtained. So preprocessing is needed for noisy grayscale brain image.

In comparison of Salt n Peppers Noisy grayscale brain image in figure 3(left) with AMF preprocessed grayscale brain image as shown figure 3(right) good image quality parameters are obtained as PSNR value is increased from 24.1446 to 56.427,MSE value reduced from 251 to 116, ER value increased from 0.93472 to 1.0377 which is near to 1, SSIM value also increased from 0.048611 to 0.94495 which is also near to 1.

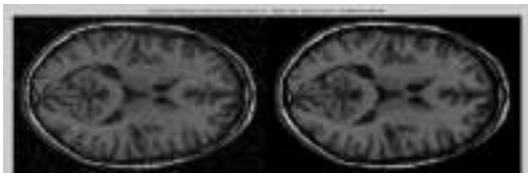


Figure 3 Salt n Peppers Noisy brain Image Vs preprocessed brain Image

Block of 8*8 matrix of preprocessed gray scale brain image as specified in table 6.1 is inputted to HWT and HWT output coefficients are obtained as shown in Table 2 in which Ngamma scaling/hard thresholding is applied and Ngamma Scaled HWT Output Block is obtained as shown in table 3. Entropy coding comprising of Zigzag scanning followed by conventional run length encoding is applied on table 3 and compressed output is obtained as shown in table 4

Block of 8x8 matrix of preprocessed gray scale brain image as specified in table 1 is inputted to HWT and HWT output coefficients are obtained as shown in Table 2 in which Ngamma scaling/hard thresholding is applied and Ngamma Scaled HWT Output Block is obtained as shown in table 3. Entropy coding comprising of Zigzag scanning followed by conventional run length encoding is applied on table 3 and compressed output is obtained as shown in table 4

Table 1 HWT Input Block brain image

97	94	94	96	96	96	96	96
97	97	98	98	98	97	96	96
96	97	97	98	98	98	96	98
89	91	94	97	98	99	99	99
83	85	89	93	96	98	99	99
79	81	83	86	89	93	96	99
73	74	75	79	86	89	94	96
66	68	72	75	79	87	93	96

Table 2 HWT Output Block brain image

382.5	389	-3.5	1	1.5	-1	0.5	0
315.25	372.25	-10.75	-13.75	-1.5	-2	-0.5	-1
3	-3.5	3	0.5	-2	-3.5	-3	-1.5
24.25	12.25	-0.75	5.25	-1.5	-3.5	-5.5	-2.5
-1.5	-3	-1.5	0	1.5	-1	-0.5	0
6.5	2	-0.5	-2	0.5	1	0.5	-1
4	6.5	6	1.5	0	-0.5	1	1.5
6.5	3.5	4.5	0.5	0.	-	2.	0.

				5	0.5	5	5

Table 3 Ngamma Scaled HWT Output Block of brain image

48	49	0	0	0	0	0	0
39	47	0	0	0	0	0	0
0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

No. of digits=11

Output after Entropy coding comprising of zigzag scan and conventional run length encoding is shown in table 3. In table 3, top left first digit is considered as DC coefficient and remaining digits are considered as AC coefficients. In zigzag scanning , number of zeros occurred between two non-zero digits are computed as length of runs. Length of runs followed by non-zero AC coefficient is considered as pair of run length. 5 pairs of run length is added with one DC coefficient and 11 digits are obtained as compressed output block of coefficients as shown in figure 4.

Table 4 Entropy coding output of Ngamma scaled HWT output Matrix

(0,49)(0,39)(1,47)(4,3)(0,0)

Run length decoding is applied the Entropy coding output in table 4 and decoded output is obtained as shown in table 5 and inverse hard thresholding followed by upscaling is applied on table 5 and output is obtained as in table 6 on which Inverse HWT is applied and decompressed output is obtained as shown in Table 7.

Table 5 Decoded Output Block of 8*8 Matrix of gray scale brain image

48	49	0	0	0	0	0	0
39	47	0	0	0	0	0	0
0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Table 6 Inverse thresholded,Upscaled decoded Output of brain image

384	392	0	0	0	0	0	0
312	376	0	0	0	0	0	0
0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Table 7 IHWT Output Block ofbrain image

96	96	96	96	98	98	98	98
96	96	96	96	98	98	98	98
96	96	96	96	98	98	98	98
96	96	96	96	98	98	98	98
84	84	84	84	94	94	94	94
84	84	84	84	94	94	94	94
72	72	72	72	94	94	94	94
72	72	72	72	94	94	94	94

After compression and decompression, degraded image is obtained. So still post processing is needed for degraded image. AMF is applied on the degraded image and AMF postprocessed image is obtained as shown in figure 4(right). In comparison of Salt n Peppers Noisy grayscale brain image in figure 4(left) with preprocessed, decompressed , AMF post processed grayscale brain image, very good image quality parameters are obtained as PSNR value is increased from 24.1446 to 76.4047,MSE value reduced from 251 to 77, ER value increased from 0.93472 to 1.0546 which is near to 1, SSIM value also increased from 0.048611 to 0.8746 which is also near to 1.

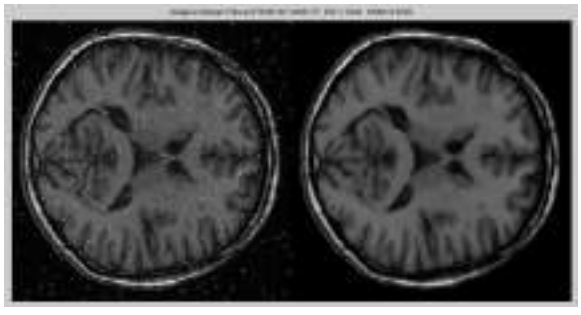


Figure 4 Salt n Peppers Noisy brain Image Vs decompressed,postprocessed brain Image

Similarly , this Compression-Decompression algorithm applied for color brain image also and output is obtained as shown in figures 5,6,7. Block of 8x8 matrix of preprocessed color brain inputted to HWT and HWT output coefficients are obtained on which Ngamma scaling/hard

thresholding is applied and Ngamma Scaled HWT Output Block is obtained then Entropy coding comprising of Zigzag scanning followed by conventional run length encoding is applied on table and compressed output is obtained. Then decompression is applied on compressed color brain image and output is analysed in table 8.

By this analysis it is observed that in the preprocessed, Decompressed, postprocessedgrayscale brain image the PSNR is highly acceptable value 76.4047 and preprocessed, decompressed, postprocessed color brain image the PSNR is highly acceptable value 74.4047 for great compression ratio 0.171875 by using scaling and hard thresholding technique with haar wavelet based compression.

Figure 5 original color brain Image Vs Salt n Peppers Noisy color brain Image

Figure 6 Salt n Peppers Noisy color brain Image Vs preprocessed color brain image

Figure 7 Salt n Peppers Noisy color brain Image Vs decompressed, post processed color brain image

Grayscale and color CT brain image are taken for the compression. In the scaling technique and hard thresholding with HWT based image compression , best Compression Ratio is achieved up to 0.171875 for both CT grayscale and color brain images with image quality parameters PSNR, MSE, ER and SSIM and shown in table 8.

Table 8 Performance parameters of grayscale and color brain image

Noise Type /Filters	PSNR	MSE	ER	SSIM
Grayscale and Color Brain Image in Medical Application				
Original grayscale brain image Vs Salt n Peppers grayscale brain image Noisy Image	24.1446	251	0.93472	0.048611
Salt n Peppers noisy grayscale brain image Vs AMF preprocessed grayscale brain image Image	56.427	116	1.0377	0.94495
Salt n Peppers grayscale brain image Noisy Image Vsdecompressed,AMFpostprocessedgrayscale brain image Image	76.4047	77	1.0546	0.8746
Original color brain image Vs Salt n Peppers noisy color brain image	24.6375	224	0.99663	0.031895
Salt n Peppers noisy color brain Image Vs AMF preprocessed color brain image	46.9018	215	1	0.95786
Salt n Peppers noisy color brain image Vs decompressed, AMF postprocessed color brain image	74.4047	197	0.89359	0.9157

V CONCLUSION

Compression ratio was estimated for grayscale and color brain image with Ngamma Scaling technique and it was proved that Ngamma Scaling with Haar wavelet Transform have given good compression ratio. While testing performance of compression and decompression algorithms in medical applications, it was ensured High compression ratio was achieved along with good visual quality of imageas compared to DCT. Adaptive median filter gives good visual quality with high PSNR value in both preprocessing and post-processing during compression and decompression of medical images. Thus this work is useful very much in fast transmission of medical images with less storage space and less bandwidth.

REFERENCES

[1] Sukhwinder Singh, Vinod Kumar, H.K.Verma ,"Adaptive threshold based block classification in medical image compression forteleradiology ",Computers in Biology and Medicine ,Vol.37, pp. 811 –819,2007

[2] J. Jyotheswar, SudiptaMahapatra ,"Efficient FPGA implementation ofDWT and modified SPIHT for lossless image compression", Journalof Systems Architecture, Vol.53, pp.369–378, 2007

[3] Gregorio Bernab , Jose M. García , José González ,"A lossy 3Dwavelet transform for high-quality compression of medical video",The Journal of Systems and Software, Vol. 82, pp.526–534,2009

[4] DCT-based subband decomposition and modified SPIHT data organization",International journal of medical informatics, Vol.76, pp. 717–725,2007

[5] R. Srikanth, A.G. Ramakrishnan, "Contextual encoding in uniform and adaptive mesh based lossless compression of MR images", IEEE Transactions on Medical Imaging,Vol. 24,pp.1199–1206,

2005.
[6] R. Shyam Sunder, C. Eswaran, N. Sriraam, "Medical image compression using 3-D Hartley transform", Comput. Biol. Med, Vol.36, pp.958–973, 2006.
[7] A.T. Deever, S.S. Hemami , "Lossless image compression with projection-based and adaptive reversible integer wavelet transforms", IEEE Transactions on ImageProcessing.vol.12 , pp.489–499. 2003.
[7] David Wu, Damian M. Tan, Marilyn Baird, John DeCampo, Chris White, and Hong Ren Wu, "Perceptually Lossless Medical Image Coding", IEEE Transactions On Medical Imaging, Vol. 25, No. 3, March 2006
[8] A. Sivanantha Raja, D. Venugopal and S. Navaneethan, "An Efficient Coloured Medical Image Compression Scheme using Curvelet Transform", European Journal of Scientific Research ISSN 1450-216X Vol.80 No.3 (2012), pp.416-422
[9] Marykutty Cyriac and Chellamuthu C., "A Novel Visually Lossless Spatial Domain Approach for Medical Image Compression", European Journal of Scientific Research ISSN 1450-216X Vol.71 No.3 (2012), pp. 347-351
[10] M.Ferni Ukrit ,A.Umamageswari ,Dr.G.R.Suresh , ―A Survey on Lossless Compression for Medical Images ‖ International Journal

An A3P approach towards Image Privacy policy recommendation on content sharing sites

Mrs. Swapna R, Dr. Nagesh H R

[1] M.Tech Student, 2 Professor ,Dept. Of CSE , Mangalore Institute of Technology and Engineering, Mangalore, Karnataka, India

***Abstract*— Popularity of the social media like Facebook, Flickr, LinkedIn and others has increasing everyday due its various features provided. Some of the features are to collaborate with friends, good and friendly user interface to share data and multimedia content at ease. Data uploaded to Social network by the user are at high risk of vulnerability. User uploaded data on the social media does play vital role in user connectivity. Thus maintaining privacy of the user uploaded data on content distribution network is at most important. To control privacy of the uploaded data user should also provide privacy information for the uploaded data. It is observed the difficulty in setting privacy for images shared by user in social network. This requires automation privacy recommendation for the user shared images based on the image metadata. This suggested privacy inference for uploaded image should suits his satisfaction level. The necessary aspect considered are social context and image metadata for analyzing the privacy setting for the user uploaded data. Two-level framework is proposed which utilizes user's history for recommending best policy setting for uploaded image. Solution provided in paper relies on an categorizing the image based on image metadata which may belong to similar policies and predicting the policy using apriori data mining algorithm to automatically suggest a policy for each image uploaded by the user in social network. The goal of this paper is to provide extensive method for privacy policy suggestions to improve the security of shared data in the social media sites.**

***Keywords*— Content sharing sites, Apriori data mining algorithm, Social media, privacy inference.**

I. INTRODUCTION

The term "Social network sites" refers to web-based services which individuals are allowed to create a user profile, and connects with list of users with whom user likes to share information, and view information of the other individuals who are already in the social network sites. Social media consumes lot of time, but still internet users are happy to use social networking sites. This will lead to the use of applications such as Facebook and Twitter becoming regular browsers, forming regular tendency that influence their daily activities in lives, both online and offline. Facebook, YouTube, Google+, Twitter and LinkedIn collectively have billions of users, and they're getting bigger and more important each and every year. Users more often utilizes online social network for establish communication with other users. Online social networks are websites which allow users to establish connections and build relationships to other Internet users. Social networking is used to be in contact with friends by building communication, make new contacts in social network of other users and find people with similar interests and ideas. Some of the tasks performed by the user are to share any special events, update user current location, share any invitation messages, current status update, and express their hobbies, wishing other person on occasion of any event and others. Most of the above task makes use of text, images or videos for sharing through the social network. The uploaded user information in social networks is stored remotely.

Information uploaded to the Social networks, stores data remotely on the server located in some other location, compare to user's personal computer. The relationship between privacy and a person's data in content sharing sites is variable. Thus it requires providing more security mechanism for the data uploaded by the user in content sharing sites. Privacy is essential to the design of security mechanisms to avoid unnecessary disclosure and violation of the privacy. Most of the content sharing site provider has offered settings to privacy. Those privacy settings can deny or allow others to access information of individual users. Individual users share their user data to the list of users depending on the content of the uploaded data. If the data uploaded to sites contains for example hobby related than data uploaded by the user to network can be shared to all in the site. If uploaded data is regarding any particular event than it will be shared with the specific group it is involved. Most content sharing sites allow individual users to set their privacy preferences. Unfortunately, some of the recent surveys have revealed that users struggle to organize and manage such privacy settings [9]. Hence, many have concluded the need of policy setting recommendation systems which can be convenient for users to configure privacy settings at ease. [2], [4], [10]. However, existing systems for automating privacy settings recommendation appear to be failing to address the individual privacy needs of images [5] considering the amount of data carried within the images, and their relationship with the online environment wherein they are exposed.

By considering required goal, we present an Adaptive Privacy Policy Prediction which focus to maintain users a settings recommendation by generating personalized policies automatically. Some of the important factors that have been considered to handle user uploaded images in A3P that influences privacy settings are:

- The privacy policy setting recommendation of user images uploaded can be maintained based on the user social activities and individual user personal characteristics.
- Metadata of the images helps in suggesting privacy settings.

II. RELATED WORK

Anna CinziaSquicciarini developed an Adaptive Privacy Policy Prediction (A3P) [1] system, a automated privacy settings recommendation for system generating personalized policies. The A3P system handles user uploaded images based on the person's personal characteristics and images metadata. The A3P system consists of two components: A3P Core and A3P Social. When a user uploads an image, the image will be first sent to the A3P-core. The A3P Core uses metadata for classification of the images and uses association rule data mining algorithm for police prediction if sufficient use data available in the system. Else if not sufficient data, A3P system will be called to fetch relevant policies. The disadvantage is inaccurate privacy policy generation in case of the absence of metadata information about the images. Also manual creation of metadata log data information leads to inaccurate classification and also violation privacy.

Jonathan Anderson proposed a tool called **Privacy Suites** [2]. This privacy suit allows users to easily choose "suites" of privacy settings. A privacy suite can be created by an expert using privacy programming. Privacy Suites could also be created directly through existing configuration UIs or exporting them to the abstract format. The privacy suite is distributed through existing distribution channels to the members of the social sites. The disadvantage of a rich programming language is less understandability for end users. Given a sufficiently high-level language and good coding practice, motivated users should be able to verify a Privacy Suite. The main goal is transparency, which is essential for convincing influential users that it is safe to use.

Alessandra Mazzia introduced PViz Comprehension Tool [5], an interface and system that corresponds more directly with how users model groups and privacy policies applied to their networks. PViz allows the user to understand the visibility of her profile according to automatically-constructed, natural sub-groupings of friends, and at different levels of granularity. Because the user must be able to identify and distinguish automatically-constructed groups, we also address the important sub-problem of producing effective group labels. PViz is better than other current policy comprehension tools Facebook's Audience View and Custom Settings page.

Chen, Chang Proposed a system named SheepDog [3] to automatically insert photos into appropriate groups and recommend suitable tags for users on Flickr. They adopt concept detection to predict relevant concepts (tags) of a photo. A reliable system toadd photos into popular groups automatically. The system recommends suitable tags for photos and provides a user friendly interface such that users could easily select their favorite tags to attach

III. METHODOLY

The proposed A3P system consists of two main components: A3P-core and A3P-social. The overall data flow as shown in Fig 1 is described as follows. When a user uploads an image, the image will be first sent to the A3P-core. The A3P-core classifies the image and determines whether there is a need to invoke the A3P-social. In most cases, the A3P-core predicts policies for the users directly based on their historical behavior.

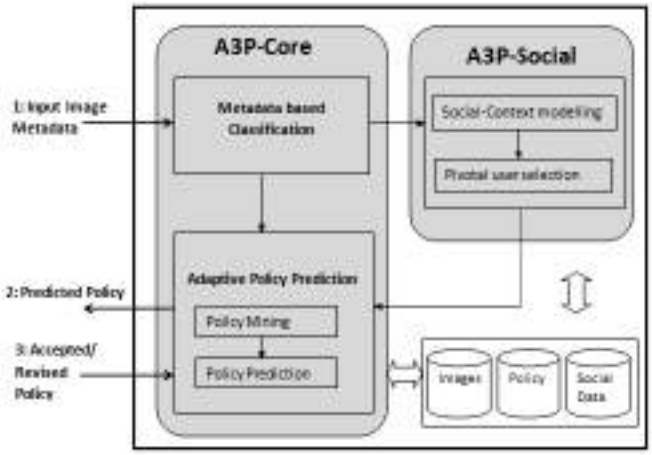


Fig.1. System Overview

A. A3P Core

The A3P Core consists of two major blocks of the framework.

- Metadata based Image Classification
- Adaptive Policy Prediction

Every image of the user gets classified based on the metadata and then its privacy policies are generalized. With the help of this approach, the policy recommendation becomes easy and more accurate. Based on the Classification based on metadata the policies are applied to the right class of images. Moreover combining the image and classification and policy prediction would enhance the system's dependency.

1) Metadata based Image Classification

Metadata can be fetched for the uploaded images in content sharing network. In general, similar images often incur similar privacy preferences.

2) Adaptive Policy Prediction

This section deals with the privacy concerns of the user by deriving the privacy policies for the images. The Adaptive Policy Prediction consists of two following sub-parts

- Policy Mining
- Policy Prediction

Policy mining deals with data mining of policies for similar categorized images and Policy prediction applies prediction algorithm to predict the policies.

- *Policy Mining*

The privacy policies are the privacy preferences expressed by the users. Policy mining deals with mining of these policies by applying different Apriori mining rules and steps. It follows the order in which a user defines a policy and decides what rights must be given to the images. This hierarchical mining approach starts by looking the popular subjects and

their popular actions in the policies and finally for conditions. It can be thoroughly reviewed with the help of following steps.

Step 1 of this process applies Apriori based data mining on the subject components of the policies of the new image. With Apriori based data mining we select the best frequent itemset.

Step 2 of this process applies a priori based data mining on the action components. Similar to the first step we will select the best frequent itemset which will give most popular combinations of action in policies.

Step 3 of this process mine the condition component in each policy set using a priori based data mining technique. The best frequent itemset are selected which gives us a set of attributes which often appear in policies.

Apriori algorithm is easy to execute and very simple, is used to mine all frequent itemsets in database. The algorithm makes many searches in database to find frequent itemsets where k-itemsets are used to generate k+1-itemsets. Each k-itemset must be greater than or equal to minimum support threshold to be frequency. Otherwise, it is called candidate itemsets. In the first, the algorithm scan database to find frequency of 1-itemsets that contains only one item by counting each item in database. The frequency of 1-itemsets is used to find the itemsets in 2-itemsets which in turn is used to find 3-itemsets and so on until there are not any more k-itemsets.

Some of the drawbacks of Association rule algorithm is that for each candidate itemset, there are as many entries as its support value that results in unnecessarily generating and counting too many candidate itemsets that turn out to be small. Apriori is used through reducing the time consumed in transactions scanning for candidate itemsets by reducing the number of transactions to be scanned. The Apriori algorithm takes advantage of the fact that any subset of a frequent itemset is also a frequent itemset. The algorithm can therefore, reduce the number of candidates being considered by only exploring the itemsets whose support count is greater than the minimum support count. All infrequent itemsets can be pruned if it has an infrequent subset. Apriori algorithm are more efficient and less time consuming.

- *Policy Prediction*

The policy mining phase may give us many policies but system needs to show the best one to the user. Thus, this approach is used to choose the best policy for the user by obtaining the strictness level. The Strictness level decides how "strict" a policy is by returning an integer value. This value should be minimum to attain high strictness. The strictness can be discovered by two metrics: a major level and coverage rate. The major level is determined with the help of combinations of subject and action in a policy and coverage rate is determined using the condition statement. Different integer values are assigned according to the strictness to the combinations and if the data has multiple combinations we will select the lowest one. Coverage rate provides a fine-grained strictness level which adjusts the obtained major level. For example a user has to 5 friends and two of them are females. Hence if he specifies policy as "friends"=male, then the coverage rate can be calculated as (3/5) =0.6. Hence, the image is less restricted if the coverage rate value is high.

B. A3P Social

The A3P-social employs a multi-criteria inference mechanism that generates representative policies by leveraging key information related to the user's social context and his general attitude toward privacy. As mentioned earlier, A3Psocial will be invoked by the A3P-core in two scenarios. One is when the user is a newbie of a site. The other is when the system notices significant changes of privacy trend in the user's social circle. In what follows, first presents the types of social context considered by A3P-Social, and then present the policy recommendation process.

1) Modeling Social Context

Social context modeling algorithm that can capture the common social elements of users and identify communities formed by the users with similar privacy concerns. The identified communities who have a rich set of images can then serve as the base of subsequent policy recommendation.

2) Identifying Social Group

The policy recommendation process is based on the social groups obtained from the previous step. Suppose User U uploaded a new image. The A3P-social will find the social group which is most similar to user U and then choose the representative user in the social group along with his images to be sent to the A3P-Core policy prediction module to generate the recommended policy for user

C. Image metadata

User gives Image metadata as input to the A3p System. This image metadata will be stored in the database for further analysis.

D. Policy

After analyzing the policy from policy prediction using image metadata, policy will further be stored in database. This helps to retrieve this policy for future analysis of the policy for other metadata uploaded by the user.

DI. Social Data

Social Data is referred to user personal information and social relationship of the user in the social network sites. This is data is used in A3P Social method to form user communities based on social context attribute and Social connection in social networking sites.

V. CONCLUSION

The proposed system for suggesting privacy setting for uploaded images helps user to set privacy at ease. In A3P, the user uploads an image which passes through a classification engine. The engine retrieves similar images based on their

content. Images that have tags falling in the same group are considered similar. Adaptive policy setting works on the premise that users set similar policies for similar photos. Once the similar images are obtained, we use a policy mining algorithm based on A priori rule mining to predict the most likely policy, which is presented to the user for approval, and modification, if necessary. The A3P system helps to mine the policies based on user available history to infer privacy policy for uploaded images. Even the social context attributes are considered during policy evaluation in case of user history unavailability.

ACKNOWLEDGEMENT

I am very thankful to my guide Dr. Nagesh H R Head of the Department, Department of Computer Science and Engineering, Mite for his cordial support, valuable information and constructive suggestions during the planning and development of this work.

REFERENCES

[1] Anna CinziaSquicciarini, Member, IEEE, Dan Lin, SmithaSundareswaran, and Joshua Wede, "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, NO. 1, January 2015.

[2] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared privacy for social networks," in Proc. Symp. Usable Privacy Security, 2009.

[3] A. Kapadia, F. Adu-Oppong, C. K. Gardiner, and P. P. Tsang, "Social circles: Tackling privacy in social networks," in Proc. Symp. sable Privacy Security, 2008.

[4] 4. KambizGhazinour, Stan Matwin and Marina Sokolova, "Yourprivacyprotector: A Recommender System For Privacy Settings In Social Networks", International Journal of Security, Privacy and Trust Management (IJSPTM) Vol 2, No 4, August 2013.

[5] 5. Alessandra Mazzia Kristen LeFevre and Eytan Adar, The PViz Comprehension Tool for Social Network Privacy Settings, Tech. rep., University of Michigan, 2011.

[6] 6. Peter F. Klemperer, Yuan Liang, Michelle L. Mazurek, "Tag, You Can See It! Using Tags for Access Control in Photo Sharing", Conference on Human Factors in Computing Systems, May 2012.

[7] 7. C. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing access control to online photo albums based on tags and linked data," in Proc. Soc. Semantic Web: Where Web 2.0 Meets Web 3.0 at the AAAI Symp., 2009, pp.

[8] 8. SergejZerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova , I Know What You Did Last Summer!:Privacy-Aware Image Classification and Search , Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012.

[9] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in Proc. 6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006, pp. 36–58.

[10] 13. K. Strater and H. Lipford, "Strategies and struggles with privacy in an online social networking community," in Proc. Brit. Comput.

Analyzing the Real Photo Images with the Scanned Photo Images using Histogram Equalizer

Dr.M.Malathy[1], Mr.Vijayanand[2], Dr.Arputha Vijaya Selvi[3]
[1]Professor, Dept. of CSE, RRCE, [2]Associate Prof., Dept. of CSE, ACSCE, VTU, Bangalore-74. [3]Dean R&D, KCE,TamilNadu.
[1]anandanmalathy@gmail.com, [2]ksgvanand@gmail.com, [3]vijayas_02@yahoo.co.in

ABSTRACT- In the biometric world, image analysis places the important role. The real acquired images are forged by the scanned photo images. Unauthorized malicious user may try to use the authorized person's scanned photo images for the access process. An authentication system should analysis difference between the real images and the scanned photo images. This paper is mainly focused on analyzing the real acquiring images with scanned photo images using histogram equalizer.

Keywords: Real acquired images, Scanned photo images, Histogram Equalizer.

I. INTRODUCTION

Recent security world, Biometric authentication system is facing the treats, vulnerabilities by the attackers. In generic biometric system has attacked by the attacker in the acquiring level, preprocessing level, feature level, database storage level, matching level and communication level. Figure 1.1 describes the basic concept of biometric attacks in various levels. Spoofing the images using synthesized images is one point of view. Synthesizing images is the reverse process of the recognition or authentication system. Analyzing the images is the forward process of the recognizing or authentication system. Sometime printed images are used to forge the authentication system. Analyzing the image in the acquiring level is place the important role to detect the unauthorized access in the access level itself. Acquiring devices like camera, sensors are not detecting the forge images. Hardware level attacking may be done in the acquiring level itself. So this image analysis is mainly focused on the acquired image is real or forged image.

The biometric authentication process compares a registered or enrolled biometric sample i.e. biometric template or identifier against newly captured biometric images.

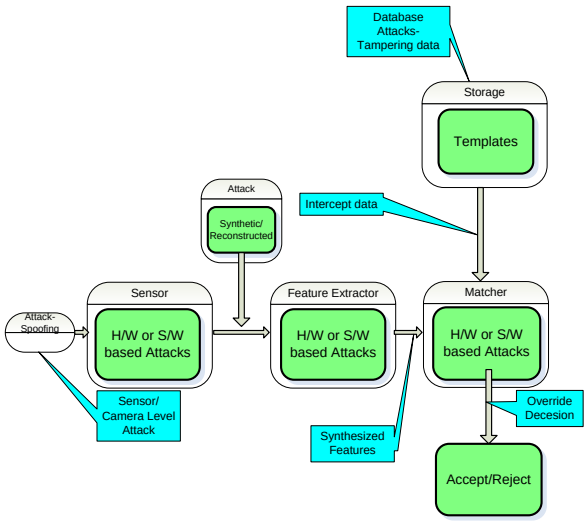


Figure 1.1 Biometric Attacks

Here, the user enters their identity to the biometric system. While entering the individual biometric traits like face, the vulnerabilities such as spoofing, collusion and coercion are occurred. The biometric system contains sensor/matcher limitations and the individuality of the biometric traits as intrinsic failure. The entered biometric traits are stored in the biometric database to provide better application/ services. The individual biometric traits are affected by the enrollment of fraud steal and made modification in the biometric images. Different types of vulnerabilities are occurred while performing this type of modification done in the enrollment.

II. RELATED WORKS

Malathy M & Arputha V.Selvi [1] have proposed, the Spoofed Iris Recognition: Synthesis of Gabor and LBP descriptor using SPPC, from this paper, the robustness of iris recognition system with spoofing attack is explained. This algorithm is a combination of Gabor wavelet followed by local binary pattern description (LBP) where the magnitude coefficient from Gabor wavelets takes as its input. Both dataset iris images and spoofed iris images are assessed by the algorithm in order to rise a genuine acceptance ratio (GAR)[1].Malathy M & Arputha V.Selvi [2] have proposed, 2-DWT and AES: Secure Authentication Management for Polar Iris Templates Using Visual Cryptography, from this paper, to protect the iris template against the spoofing attacks in the database storage level. Two shares were stored separately and merging the shares then only authentication system accessed the genuine user. Malathy M & Arputha V.Selvi [3] have proposed, the liviness face detection based on the binary image of the eye images. These eye images are cropped from the face images and photo face images, the gray scale value of the photo eye image had converted in to binary images and found the liviness. Akhtar, et al [4] have investigated, a real spoof attack samples that verify the multimodal biometric systems. Spoofed face and iris samples were replicated with a photo attack method. The photo of each individual was put in front of the capture device. While spoofed fingerprint samples was created by the same method. For each individual, 10 spoofed face, fingerprint and iris samples were created. The biometric systems were not intrinsically robust against spoof attacks contrary to the common belief. It can be cracked by spoofing only one biometric trait. Schwartz, et al[5] have presented a

face spoofing detection through partial least squares and low-level descriptors. Partial Least Squares regression to provide a feature weighting to distinguish between live and spoof images or videos. The use of a robust set of feature descriptors renders many classical machine learning methods intractable due to extremely large resulting feature space, which becomes more evident when the temporal function was considered. The reduced number of training samples was compared to the number of descriptors.

III. PROPOSED IMAGE ANALYZER

In the proposed Image analyzer, the real face image and the scanned photo image were analyzed according to the variations in the histogram equalizer. The Fig 1.2 shows the system architecture of the image analyzer. It has two modules. One is acquiring the face image in the different type of devices. The face image captured from the webcam and stored it in the system. The photo face image scanned from the scanner and stored it in the system. The both images are given to the input of the image analyzer. The analyzer has to find the contrast of images using the histogram equalizer. The histogram equalizer, the image is divided in to two frames horizontally vertically row and col of the pixel intensity taken to find the cumulative sum of the image. The contrast, correlation, energy and homogeneity are found by the image analyzer based on the following equations, Eq.No.(1) represents the mathematical equation of the contrast.

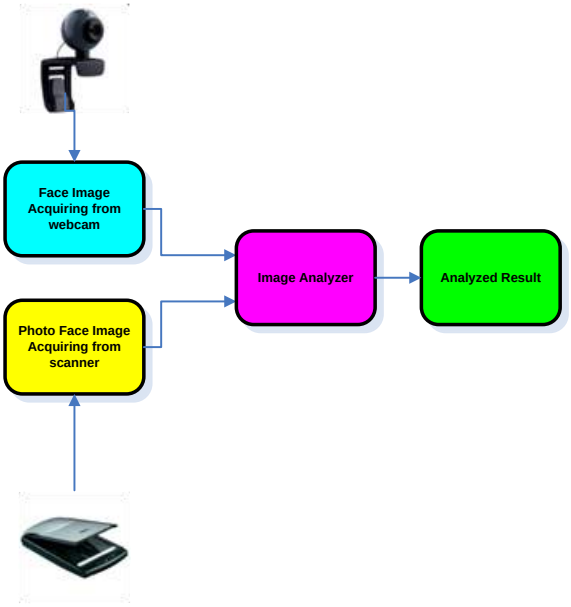


Fig.1.2 System Architecture of Image Analyzer

$$\text{Contrast} = \sum_{(i,j)} [i-j]^2 P(i,j) \quad \ldots\ldots\ldots\ldots \quad (1)$$

Where p=image, i,j=coordinates , p(i.j)=Intensity value at i,j

Eq.No.(2) represents the mathematical equation of the correlation.

$$\text{Correlation} = \sum_{(i,j)} P(i,j) \left[\frac{(i-\mu)(i-\mu)}{\sqrt{(\sigma)2(\sigma)2}}\right] \quad \ldots\ldots\ldots \quad (2)$$

Eq.No.(3) represents the mathematical equation of the Energy.

$$\text{Energy} = \sum_{(i,j)} P(i,j)^2 \quad \ldots\ldots\ldots\ldots \quad (3)$$

Eq.No.(4) represents the mathematical equation of the Homogeneity.

$$\text{Homogeneity} = \sum_{(i,j)} \frac{P(i,j)2}{1+(i-j)} \quad \ldots\ldots\ldots \quad (4)$$

The final result, the differences of the values of all the parameters are found and using these different values analysis shown the original and fake images are found for an authentication system.

IV. IMPLEMENTATION RESULT

The implementation results are given below. The Fig. 1.3 shows the Original webcam image that corresponding histogram image and also histogram equalized image with the corresponding histogram equalized image.

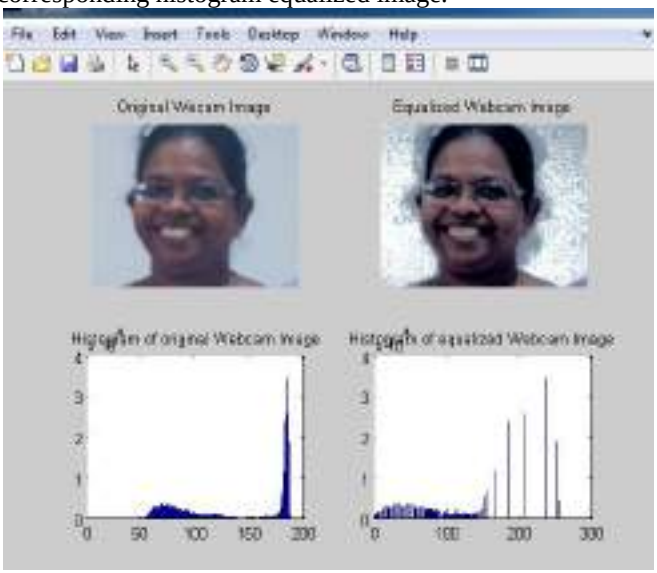


Fig. 1.3 Webcam image with the corresponding histograms

The Fig. 1.4 shows the Scanned photo image that corresponding histogram and also histogram equalized scanned photo image with the corresponding histogram equalized image.

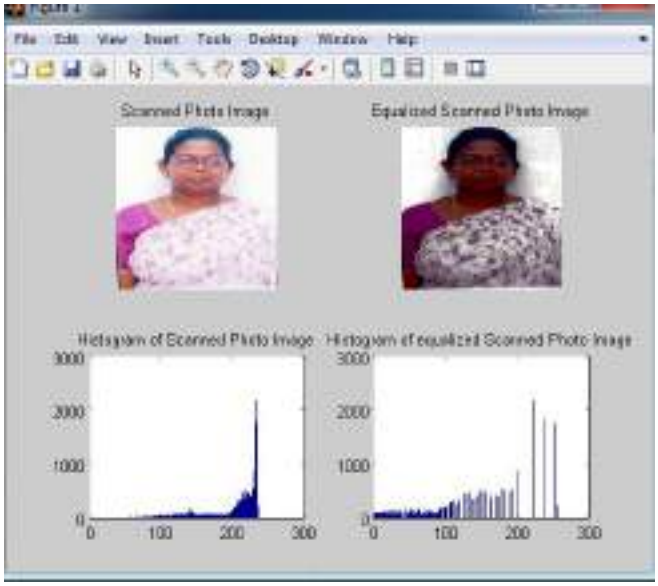


Fig. 1.4 Scanned photo image with the corresponding histograms

Table 1.1 Difference between the parameters of scanned photo image and the real photo image

Sl. No	Type of the Images	Images	Contrast	Correlation	Energy	Homogeneity
1	Scanned Photo Image		**0.239**	**0.945**	**0.198**	**0.907**
2	Real Photo Image		**0.036**	**0.992**	**0.321**	**0.982**
Difference =(Real Photo image - Scanned photo image)			**-0.203**	**0.047**	**0.123**	**0.075**

The table 1.1 shows the different between the parameters of the two types of images.

V. CONCLUSION

Comparison of two face images, one is webcam face image, another one is scanned face image, both inputs are given to the image analyzer, the image analyzer analysis the images based on the histogram equalizer. Finally, conclude that the parameters differences are, contrast is -0.203, correlation is 0.047, energy is 0.123 and homogeneity is 0.075. Based on the results the two images are different. It will useful to identify the original and fake images in the authentication system. In future work, database should be increase and train the image analyzer to find the printed photo image, all type of original and fake biometric images.

REFERENCES

[1] Malathy, M & Arputha Vijaya Selvi,J,'Spoofed Iris Recognition: Synthesis of Gabor and LBP descriptor using SPPC', Australian Journal of Basic and Applied Sciences(ISSN 1991-8178), pp. 433-442. 2014

[2] Malathy, M. & Arputha Vijaya Selvi, J., "2-DWT and AES: Secure Authentication Management for Polar. Iris Templates Using Visual Cryptography," Journal of Testing and Evaluation, Vol. 45, No. 2, 2017, pp. 1–2015; published online February 2, 2016.

[3] Malathy, M & Arputha Vijaya Selvi,J,'Face Liveness Authentication/Anti-Spoofing Engine using Morphological- shared weight Neural Network', Advances in Mathematics Scientific Developments and Engineering Application,Narosa Publishing House Pvt.Ltd., ., ISBN-978-81-8487-074-9, pp.550-557. 2010

[4] Akhtar, Zahid, and Sandeep Kale. "Security Analysis of Multimodal Biometric Systems against Spoof Attacks." Advances in Computing and Communications. Springer Berlin Heidelberg, 2011. 604-611.

[5] Schwartz, William Robson, Anderson Rocha, and Helio P. Edrini. "Face spoofing detection through partial least squares and low-level descriptors."Biometrics (IJCB), 2011 International Joint Conference on. IEEE, 2011.

[6] Heo, Jingu, and Marios Savvides. "Gender and ethnicity specific generic elastic models from a single 2D image for novel 2D pose face synthesis and recognition." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.12 (2012): 2341-2350.

[7] Malathy M & Arputha Vijaya Selvi J, "Survey of Cyber Security and Laws for Social Media Sites (SMS)', Proceedings of International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS-2016) at Kings College of Engineering, Thanjavur India, 978-1-4673-6725-7/16/213-216 ©2016 IEEE

[8] Malathy, M & Arputha Vijaya Selvi,J 2012,'Multimodal Biometric Decision Fusion for Liveness Authentication Authentication / Anti-Spoofing Engine', Electronic Design and Signal Processing Narosa Publishing House Pvt.Ltd., ISBN 978-81-8487-160-9, pp.195-201.

[9] Feng, Zhen-Hua, et al. "Random cascaded-regression copse for robust facial landmark detection", IEEE Signal Processing Letters 1.22 (2015): 76-80.

[10] M. Malathy & J. Arputha Vijaya Selvi,J 2011," Multimodal Biometric Expert Decision Fusion", International Journal of Mathematics, Computer Sciences and Information Technology Vol. 4, No. 1, January-June 2011, pp. 113-123.

A Survey on Dynamic Web Services Composition Existing Methods and Related Techniques

Neha Singhal, Dr.Y.S Kumaraswami
Assistant professor, Department of Information science and Engineering, Rajarajeswari College of Engineering, Bangalore
Professor, Department of Information Science and Engineering, Nagarjuna College of Engineering & Technology, Bangalore
neha.banasthali@yahoo.co.in, Ysksldswami2@yahoo.com

ABSTRACT- In recent time, as technology is the basic need and web service is emerging as the solution for many problems. Web is the biggest source of information and services. Most of the services are already in existence. The web service composition means to develop new service by composing the related existing services. Optimize Web service selection is still a challenging task. In this paper, we present our observation and a technical survey on 'Dynamic Web Services selection and Composition Techniques' after a long term study and review various papers.

Keywords: Web Service, Web Services Composition

I. INTRODUCTION

In this Computerized Era, approximately everything, from shopping to travelling, is available on the web. Web Service are scattered over the network. The nature of web services is heterogeneous and dynamic [1]. Platform and language dependency is not the constraint for web services [1][2]. Web services can be created and updated very dynamically. These available web services can be used to improve the business performance by reusing the available web services. To increase the business performance we need to compose these heterogeneous, platform and language independent services together. For that manual composition is not a solution so we want an computerized composition method. In this paper we are performing a survey on existing automated web service composition methods and analysing their performance. Web service is a self contained, modular application which is located and invoked over the internet. A web service can be considered as a module which can be identified by a unique URL. To enhance the performance of B2B collaboration inside and outside the boundary of the organization we need to compose these available web services.

Web Services can be categorized as -

Standalone Web Service: Single functionality services are called as standalone services.

Composite Web Service: complex service and a combination of various functionalities and combination of various services is called as composite web services.

Stand alone Web Service

Composite Web Service

Fig. 1

II. INTRODUCTORY APPROACH TO WEB SERVICE COMPOSITION

In traditional approach of Web Service Composition, the services are gathered and then combined manually based on the particular requirement. To do so, two approaches were taken into consideration- Top Down and bottom up Approach. In Top –down Web Service composition approach, first we find out a workflow, and then we find out the Services which can be added to complete the workflow. In bottom-up approach, we collect all the related Web Services and then define the workflow consisting of all the services to achieve the desired functionality. Both the approaches involve a lot of manual effort which is time consuming as well. To overcome the effort and make it faster, automation approach is required for composition of services. However, creating fully automatic system is an issue with bunch of quality of service constraints.

III. CHALLENGES OF AUTOMATIC COMPOSITION

Web services are heterogeneous in nature. Though they have a very big advantage of being language platform independent, but due to their scattered nature, it is very challenging to compose these services. In fully automated approach, the manual effort is needed for defining starting and ending points and certain criteria. However, the approach addresses the functional requirement of system and non-functional requirements are not taken into consideration. Discovering the non-functional aspects of the solution could be really challenging task, but it can not be ignored. In a long term, these non-functional aspects can result into complex issues.

There is huge number of web services with the similar functionality. Those having similar functionality are

distinguishable via QoS values. Several QoS factors are as follows:-

Throughput: - Throughput is known as the business rate able to produce a product for a given unit of time.

Response time: - Response time is the diversity between time takes place in respond and request of service

Cost: - Cost is usually a budgetary calculation of time, resources, material, effort and utilities consumed, risks incurred, and opportunity in production and delivery of a good or service

Reliability: - Reliability in the broad meaning in the science aimed at forecast, analysing, prohibit and assuage failures over time.

Network delay: - The delay of a network describes about what time it used for a piece of data to transfer across the network from one source or endpoint to another. It is normally calculates in multiples or parts of seconds.

Reputation: - Reputation of a social entity (a person, a social group, an organization) is an assumption about that entity, normally a result of social evaluation on a set of criteria.

IV. RESEARCH REVIEW FOR AUTOMATED COMPOSITION

Web service composition can be considered of four types:-

- Semantic based
- Model driven based
- Ontology based
- Context based

Summary of web service composition approaches:-

Features of web service languages:-

Web service composition is an important element of service oriented architecture.web service composition provides the reusability benefits. Overview is presented in various web service composition languages namely BPML, BPEL4WS, OIL, DAML and WS-CDL.

BPML: - The Business Process Management Initiative (BPMI) has formed a standard BPMN as Business Process Modelling Notation. The BPMN 1.0 defines was liberated released in May, 2004. This specification shows the result of performance of BPMI Notation working group. The main objective of the BPMN effort is to give notation which is understandable by all business users, from the business analysts that define the starting outline of the processes, to the technical developers important for implementing the technology that will perform those processes, and at last, to the business people who will regulate and supervisor those processes. BPMN will also be supported with a private model that will facilitate the generation of executable BPEL4WS. Thus, BPMN adds a levelness bridge for the distance between the business process design and process fulfilment. BPMN specifies a Business Process Diagram (BPD), which is based on a pictorial representation technique tailored for creating pictorial sketch process of business operations.

Approach	Semantic based	Model driven	Ontology driven	Context based
Context support	no	no	no	yes
Execution monitoring	no	no	no	yes
Semantic support	yes	no	yes	no
Transaction support	no	yes	no	no
QoS monitoring	no	yes	yes	yes
Coordination	no	yes	yes	no

Table :-1

BPEL4WS:- Today we use web services for communication with each other, announce them, and are exposed and request using wide specifications of industry. However, until last week, associating these services together into a business scheme or a combination gave the user number of clashing specifications to choose from as was the option with WSFL from IBM and XLANG from Microsoft. BPEL4WS as The Business Process Execution Language for Web Services represents the combination of WSFL and XLANG, and with profit, will become the basic of a common for web service composition. BPEL4WS combines the best of both WSFL (groundwork for graph adapted processes) and XLANG (skeletal constructs for processes) into one adhesive package that supports the application of any kind of business process in a very basic pattern. In addition to being an application language, BPEL4WS can be used to explain the interfaces of

business processes as well using the notation of basic processes. We will describe further on this in future articles.

OIL: - This document explained the OWL Web Ontology Language. OWL is basically to be used when the information of documents wants to be processed by any operation, as equivalent to situations where the data only wants to be show to humans. OWL can be used to clearly show the meaning of terms in meanings and the relationships between those terms. This shows the terms and their interrelationships is called ontology. OWL has more facilities for expressing explanation and interpretation than XML, RDF, and RDF-S, and thus OWL goes behind these languages in its capacity to represent machine understandable content on the Web. OWL is a review of the DAML+ OIL web ontology language whole lessons learned from the design and application of DAML+OIL.

DAML:-DAML (DARPA Agent Markup Language) is an html language for the DARPA as U.S. Defence Advanced Research Project Agency that is based on the Extensible Markup Language (XML). DAML is described to have a higher capacity than XML for explaining objects and the relationships between objects, to express meaning, and to develop a higher level of property of a product or system among Web sites. As the central research and development agency for the U. S. Department of Defence, DARPA was serving as a means of pursuing an aim in the creation of the Internet and many of its technologies. DARPA is constructing DAML as a technology with the ability to acquire built into the language by the nature of agents, programs that can dynamically identify and comprise sources of information, and communicate with other agents in an autonomous fashion.

WS-CDL: -The WS Choreography Definition Language (WS-CDL) is an XML-based language that describes peer-to-peer combination of Web Services participants by defining their common and integral observable behaviour; where arranged message exchanges result in accomplishing a common business goal. The Web Services Choreography specification is aimed for composing interoperable peer-to-peer collaborations between any types of Web Service participant
Regardless of the auxiliary platform or programming system used by the implementation of the hosting environment.

Web service languages comparison:

	BPML	BPEL4 WS	OIL	DMAL	WS-CDL
Semantic support	no	no	high	high	no
Transaction Support	medium	medium	medium	medium	medium
Exception handling	high	high	high	high	medium
Collaboration support	medium	high	high	high	low
Business collaboration	no	medium	no	no	no
Software vendor support	high	high	low	low	medium
Workflow control	high	high	high	high	low
Role modeling	low	low	no	no	high

Table :-2

QoS aware web service composition approaches:-

categories	Sub-categories	techniques
Static approach	Local optimization	Dynamic programming
		DFS
		Simple additive weighing
	Linear optimization	Linear programming

	approximation	Genetic algorithm
		Particle swarm algorithm
	Pareto optimization	Genetic algorithm
		Particle swarm algorithm
		Weighing
Dynamic approach	Internal adaption	AI-Planning
		Strengthened learning
		Graph based approach
		Heuristic algorithm
	External adaption	Social network analysis

Table :-3

CONCLUSION

In this paper a survey has been performed on web service composition methods. Several QoS parameters have been discussed for the optimum web service selection. Web service composition algorithms are categorized based on their nature.

REFERENCES

[1] Jinghai Rao and Xiamen Su, "A Survey of Automated Web Service Composition Methods".

[2] Narges Hesami Rostami and Esmaeil Kheirkhah and MehrdadJalali "Web Services Composition methods and techniques: a review" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.3,No. 6,December 2013

[3] Furkh Zeshan and Radziah Mohamad "Semantic Web Service Composition Approaches: Overview and Limitations" International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(3): 640-651 The Society of Digital Information and Wireless Communications, 2011 (ISSN: 2220-9085)

[4] Narges Hesami Rostami and Esmaeil Kheirkhah and Mehrdad Jalali "An Optimized Semantic Web Service Composition Method Based on Clustering and Ant Colony Algorithm"International Journal of Web & Semantic Technology (IJWesT) Vol.5, No.1, January 2014.

[5] Tongguang Zhang "QoS-aware Web Service Selection based on Particle Swarm Optimization" journal of networks, vol. 9, no. 3, march 2014

[6] Dr. IIavarasan Egambaram and G. Vadivelou and S. Prasath Sivasubramanian "Qos based web service selection"

[7] Biplav Srivastava and Jana Koehler "Web Service Composition - Current Solutions and Open Problems"

[8] Petia Wohed and Wil M.P. van der Aalst and Marlon Dumas and Arthur H.M. ter Hofstede "Analysis of Web Services Composition Languages: The Case of BPEL4WS" Research conducted while at the Queensland University of Technology

[9] Ravi Khadka and Brahmananda Sapkota" An Evaluation of Dynamic Web Service Composition Approaches"

[10] Srivastava, B., Koehler, J."Web Service Composition - Current Solutions and Open Problems"In: ICAPS 2003 Workshop on Planning for Web Services. (2003) 28–35

[11] Dustdar, S., Schreiner, W.: A survey on web services composition. International Journal of Web and Grid Services 1 (2005) 1–30

[12] Wang Li, He Yan-xiang, "Web Service Composition based on QoS with Chaos Particle Swarm Optimization ", 2010 IEEE-978-1-4244-3709-2/10

[13] Li Ying, "A Method of Automatic Web Services Composition Based on Directed Graph", 2010 International Conference on Communications and Mobile Computing.

[14] Yu Zhang, "Web Service Composition Based on Mobile Agent", 2011 IEEE-978-0-7695-4539-4/11International Conference on ommunications and Mobile Computing.

[15] Yang Syu, "A Survey on Automated Service Composition Methods and Related Techniques", 2012 IEEE-978-0-7695-4753-4/12 Ninth International Conference on Services Computing.

[16] George Zheng and Athman Bouguettaya, "Service Mining on the Web"Ieee Transactions On Services Computing, Vol. 2, No. 1, January-March 2009

An Approach to Mining Sequential Rules Common to Several Sequences

G.Jayagopi1, Dr.S.Pushpa2

1Research Scholar, Department of CSE, St.Peter's University, Avadi, Chennai, India.

2Professor, Department of CSE, St.Peter's University, Avadi, Chennai, India.

jayagopi_sv@yahoo.in, Pushpasangar96@gmail.com

ABSTRACT-Sequential pattern mining is an important data mining task with wide applications. It consists of discovering subsequences that are common to multiple sequences.Several algorithms have been proposed for this task such as GSP, PrefixsSpan,SPADE. However , sequential patterns found by these algorithms are often misleading for the user. The reason is that patterns are found solely on the basis of their support. Sequential patterns is not straightforward because they can contain multiple items and sequential pattern mining algorithms have just not been designed for that. To mine Partially-ordered sequential rules, we propose the RuleGrowth algorithm, which is efficient and easily extendable. In particular, we present an extension (TRuleGrowth) that accepts a sliding window constraint to find rules occurring within a maximum amount of time. A Performance study with four real –life datasets show that RuleGrowth and TRuleGrowth have excellent performance and scalability compared to baseline algorithms and that the number of rules discovered can be several orders of magnitude smaller when the sliding window constraint is applied. Partially ordered sequential rules can provide a much higher prediction accuracy than regular sequential rules for sequence prediction.

Keywords: **Sequential rules, Sequential patterns, Sequence, data mining.**

I INTRODUCTION

Sequential pattern mining has focused extensively on mining symbolic sequences. A sequential pattern is a frequent subsequence existing in a single sequence or a set of sequences. A sequence α =($a_1a_2…a_n$) is a subsequence of another sequence β=($b_1b_2…b_m$) if there exist integers $1\leq j_1<j_2<…<j_n\leq m$ such that $a_1 \subseteq b_{j1}$, $a_2 \subseteq b_{j2}$,…,an $\subseteq b_{jn}$. For example, if α = ({ab},d) and β =((abc},{be},{de},a), where a,b,c,d,and e are items, then α is a subsequence of β. Mining of sequential patterns consists of mining the set of subsequences that are frequent in one sequence or a set of sequences. Many scalable algorithms have been developed as a result of extensive studies in this area. Alternatively, we can mine only the set of closed sequential patterns, where a sequential pattern s is closed if there exists no sequential pattern s', where s is a proper subsequence of s', and s' has the same (frequency) support as s. Similar to its frequent pattern mining counterpart, there are also studies on efficient mining of multidimensional, multilevel sequential patterns. As with constraint – based frequent mining, user-specified constraints can be used to reduce the search space in sequential pattern mining and derive only the patterns that are of interest to the user. Moreover, we may relax constraints or enforce additional constraints on the problem of sequential pattern mining to derive different kinds of patterns from sequence data. For example, we can enforce gap constraints so that the patterns derived contain only consecutive subsequences or subsequences with very small gaps. Alternatively, we may derive periodic sequential patterns by folding events into proper-size windows and finding recurring subsequences in these windows. Another approach derives partial order patterns by relaxing the requirement of strict sequential ordering in the mining of subsequence patterns. Besides mining partial order patterns, sequential pattern mining methodology can also be extended to mining trees, lattices, episodes, and some other ordered patterns.A sequential rule indicates that if some events occur, some other events are likely to follow with a given confidence. Sequential rule mining has been applied in several domains such as drought management, stock market analysis, weather observation, reverse engineering, e-learning and e-commerce. Algorithms for sequential rule mining are designed to either discover rules appearing in a single sequence, across sequences or common to multiple sequences. In this article, we are interested by the task of mining sequential rules common to multiple sequences, which is analogous to sequential pattern mining and is also applied on sequence databases. It consists of finding rules of the form X⇒Y in a sequence database such that X and Y are sequential patterns. Each rule is found on the basis of its support (the percentage of sequences that contains the rule) and its Confidence (the probability that the sequential pattern Y will appear after X)[1]. Those rules are interpreted as "if X appears, Y is likely to appear with a given confidence afterward". Besides, there are many other applications such as:

A sequence is an ordered list of events. Sequences may be categorized into three groups, based on the characteristics of the events they describe: (1) time-series data, (2) symbolic sequence data, and (3) biological sequences. Let's consider each type.

In **time-series data**, sequence data consist of long sequences of numeric data, recorded at equal time intervals (e.g., per minute, per hour, or per day). Time-series data can be generated by many natural and economic processes such as stock markets, and scientific, medical, or natural observations.

Symbolic sequence data consist of long sequences of event or nominal data, which typically are not observed at equal time intervals. For many such sequences, gaps (i.e., lapses between recorded events) do not matter much. Examples include customer shopping sequences and web click streams, as well as sequences of events in science and engineering and in natural and social developments.

Biological sequences include DNA and protein sequences. Such sequences are typically very long, and carry important, complicated, but hidden semantic meaning. Here, gaps are usually important.

II. BACKGROUND

A **sequence database** [2] SD is a set of sequences S={s1,s2,…sm} and a set of **items** I={*i*1, *i*2,…*i*n} occurring in these sequences, where each sequence is assigned a unique **SID** (Sequence ID). A **sequence** is an ordered list of **itemsets** (sets of items) *sx*=I1, I2, … Ip such that I1, I2, …Ip ⊆ I. For instance, Table 1 depicts a sequence database containing four sequences respectively having the sids *seq1*, *seq2*, *seq3* and *seq4*. In this example, each single letter represents an item. Item(s) between curly brackets represent an itemset. For instance, the sequence *seq1* means that items *a* and *b* occurred at the same time, and were followed successively by *c*, *f*, *g and e*.sequences, where each sequence is assigned a unique **SID** (Sequence ID). A **sequence** is an ordered list of **itemsets** (sets of items)

TABLE 1

A SEQUENCE DATABASE

SID	Sequences
Seq1	{a,b},{c},{f},{g},{e}
Seq2	{a,d},{c},{b},{a,b,e,f}
Seq3	{a},{b},{f},{e}
Seq4	{b},{f,g,h}

A sequence where each itemsets is annotated with a timestamp is called a **time-sequence**. For example, Fig. 1 shows a time-sequence containing 13 itemsets. A **window** is a group of consecutive itemsets in a sequence. A **slid-ing-window** is a window that is assumed to move from the beginning of a sequence to its end, one itemset (or one time unit) at a time. For example, a sliding-window with a length of 3 time units can move in 15 different positions *w1*, *w2*, ... *w15* for the sequence depicted in Fig. 1 (each position is said to be a window). Note that some windows (*w1*, *w2*, *w14*, *w15*) extend outside the sequence so that each itemset appears in the same number of windows.

III EXISTING SYSTEMS

In previous work, we proposed two algorithms named CMRules and CMDeo for mining partially-ordered sequential rules, which will be used as baseline algo- rithms in this article.

CMRules is based on the idea that partially-ordered sequential rules can be seen as a subset of association rules. CMRules performs two steps to discover se- quential rules.

First, it ignores the temporal information from the sequence database taken as input to mine asso- ciation rules [1]. Then, to obtain sequential rules from association rules, CMRules scans the original database to eliminate rules that do not meet *minsup* and *minconf* ac- cording to the sequential ordering. The main benefits of CMRules are that association rule mining algorithms can be reused to implement the algorithm and that it per- forms better than CMDeo for some datasets. Its main drawback is that its performance depends on the number of association rules. If this set is large, CMRules becomes inefficient.

CMDeo proceeds by first scanning the database to generate rules of size 1*1 (containing two items). The algorithm then recursively finds larger candidate rules by combining frequent rules of smaller size in a level-wise manner (similar to Apriori). This is done by two sepa- rate processes. Left-side expansion is the process of taking two frequent rules X⇒Y and Z⇒Y, where X and Z are itemsets of size *n* sharing *n*-1 items, to generate a larger candidate rule X∪Z⇒Y. Right-side expansion is the process of taking two frequent rules Y⇒X and Y⇒Z, where X and Z are itemsets of size *n* sharing *n*-1 items, to generate a larger candidate rule Y⇒X∪Z. After candidate rules are generated, their support and confidence are calculated by scanning original sequences of the database. To prune the search space of candidate rules, CMDeo uses the property that expanding the left side of a rule not respecting *minsup* will not result in valid sequential rules, and that expanding the right side of a rule not respecting *minsup* or *minconf* will not generate valid sequential rules . In ,we extensively compared the performance of CMDeo and CMRules. It was found that CMDeo performs considerably better than CMRules for some datasets. But for others, the search space is such that CMDeo generates a very large number of candidate rules that are invalid, which makes CMRules more efficient.

IV PROPOSED SYSTEMS

The RuleGrowth algorithm that we propose in this arti- cle avoids this problem of candidate generation by in- stead relying on a pattern-growth approach partly in- spired by the one used in the PrefixSpan algorithm [14] for sequential pattern mining[2]. RuleGrowth first finds rules of size 1*1 and then recursively grows them by scanning the sequences containing them to find single items that can expand their left or right sides. This strate- gy ensures that only rules occurring in the database are considered as potential valid rules by the algorithm. We name the two processes for expanding rules in RuleGrowth *left expansion* and *right expansion* akin to the homonym processes of CMDeo. Note however that these processes are different; RuleGrowth finds larger rules by adding items to rules by scanning sequences containing the rules (a depth-first search), whereas CMDeo combine pairs of rules to generate candidates (a breadth-first search). Another distinctive feature of RuleGrowth is that it keeps track of the first and last occurrences of each item, and also of antecedents and consequents to avoid scanning sequences completely.

RULEGROWTH(S, *minsup*, *minconf*)

1. **Scan** the database S once. For each item *c* found, record the *sids* of the sequences that contains *c* in a variable *sids*(*c*).
2. **FOR** each pair of items *i*, *j* such that | *sids*(*i*) | / |S| = *minsup*
3. and |*sids*(*j*)| / |S| = *minsup* {
4. *sids*(*i*⇒*j*) := Ø. *sids*(*j*⇒*i*) := Ø.
5. **FOR** each sid *s* ∈ (*sids*(*i*) n *sids*(*j*)) {

6. **IF** *i* occurs before *j* in *s*, *sids*(*i*⇒*j*) := *sids*(*i*⇒*j*) ∪ {*s*}.
7. **IF** *j* occurs before *i* in *s*, *sids*(*j*⇒*i*) := *sids*(*j*⇒ *i*) ∪ {*s*}. }
8. **IF** (|*sids*(*i*⇒*j*)| / |S|) = *minsup* **THEN {**
9. **EXPANDLEFT**({*i*}⇒{*j*}, *sids*(*i*), *sids*(*i*⇒*j*)).
10. **EXPANDRIGHT**({*i*}⇒{*j*}, *sids*(*i*), *sids*(*j*), *sids*(*i*⇒*j*)).
11. **IF** (|*sids*(*i*⇒*j*)| / |*sids*(*i*)|) = *minconf* **THEN OUTPUT** rule {*i*}⇒{*j*} with its confidence and support. **}**
12. **IF** (|*sids*(*j*⇒*i*)| / |S|) = *minsup* **THEN {**
13. **EXPANDLEFT**({*j*}⇒{*i*}, *sids*(*j*), *sids*(*j*⇒*i*)).
14. **EXPANDRIGHT**({*j*}⇒{*i*}, *sids*(*j*), *sids*(*i*), *sids*(*j*⇒*i*)).
15. **IF** (|*sids*(*j*⇒*i*)| / |*sids*(*j*)|) = *minconf* **THEN OUTPUT**
16. rule {*j*}⇒{*i*} with its confidence and support. }}

Fig. 2. The RuleGrowth algorithm

V EXTENDING RULEGROWTH TO DISCOVER RULES WITH A SLIDING-WINDOW

Because RuleGrowth grows rules one item at a time, constraints can be easily added to the algorithm for the needs of specific applications. For example, it would be easy to add constraints on the number of items that rules can contain or to restrict the items that are added to rules.In this section, we present one particular extension which is to discover rules occurring within a sliding- window, i.e. within a maximum number of consecutive itemsets in each sequence. We present this extension be- cause applying a sliding-window has shown to be very useful for the discovery of temporal patterns for many real-life applications such as analyzing sensor networks and stock market data, because users often only wish to discover patterns occurring within a maximum amount of time . For this reason, several data mining algorithms use a sliding-window or have been modified to accept one we name TRuleGrowth the extension of RuleGrowth that discovers rules while verifying that they occur in a sliding-window. As it will be shown, discovering rules appearing in a sliding-window has several important benefits. First, it can decrease the execution time by several orders of mag-nitude by pruning the search space. Second, it can pro- duce a much smaller set of rules, thus reducing the disk space requirement for storing rules found and making it easier for the user to analyze results. Third, setting a window constraint can increase prediction accuracy when rules are used for prediction. Note that adding a sliding-window could also be done with CMRules/CMDeo. However, it is best done with RuleGrowth because it can check the window constraint when it scans sequences to search for items, whereas CMRules/CMDeo can only verify that rules respect the the window after rules have been generated (they would generate many rules not respecting the sliding-window).

First, instead of keeping the first and last occurrences of each item for each sequence all occurrences of each item are now kept for each sequence. Recall that an occurrence of an item for a sequence is represented as an integer indicating the position of an Itemset containing the item.

For example, the occurrences of *a* in sequence *seq2* = {*a*, *d*},{*c*},{*b*},{*a*, *b*, *e*, *f*} are 1 and 4 (*appears* in the first itemset and the fourth itemset), and the occurrences of *b* in *seq2* are 3 and 4 (*b* appears in the third itemset and the fourth itemset). The second change is to modify line 6 and line 7 of the RULEGROWTH procedure so that when checking if item *i* occurs before item *j* and if *j* occurs before *i* in a sequence, the check also verifies that it is true within *window_size* consecutive itemsets. This check is performed efficiently by comparing each occurrence of *i* with each occurrence of *j* for the sequence by using the hash tables. If there exists an occurrence *x* of *i* and an occurrence *y* of *j* such that $y - x > 0$ and $y - x + 1 = window\_size$, then it is concluded that *i* occurs before *j* in the sequence while respecting the sliding-window. Similarly, if there exists an occurrence *x* of *j* and an occurrence *y* of *i* such that $y - x > 0$ and $y - x + 1 = window\_size$, then it is concluded that *j*occurs before *i* in the sequence while respecting the sliding-window. For example, consider items *a* and *b* in sequence *seq2*={*a*, *d*},{*c*},{*b*},{*a*, *b*, *e*, *f*} and *window_size* = 3. By comparing occurrences of *a* and *b*, TRuleGrowth finds that item *a* appears before *b* while respecting the slidingwindow because for the occurrence 1 of *a* and the occurrence 3 of *b*, $3 - 1 > 0$ and $3 - 1 + 1 = 3 = window\_size$. The algorithm will also discover that *b* appears before *a* while respecting the sliding-window because $4 - 3 > 0$ and $4 - 3 + 1 = 2 = window\_size$.

VI PERFORMANCE EVALUATION

To evaluate Rule Growth and TRuleGrowth, we compared their performance with CMRules and CMDeo. Experiments were performed on a notebook computer with a 2.53 Ghz P8700 Core 2 Duo processor running Windows XP and 1 GB of free RAM. Algorithms were implemented in Java.

CONCLUSION

This paper presented two algorithms. Rule Growth is a novel algorithm for mining sequential rules common to multiple sequences. Unlike previous algorithms, it uses a pattern-growth approach for discovering valid rules such that it avoids considering rules not appearing in the database. The second algorithm (TRuleGrowth) allows the user to specify a sliding-window constraint on rules to be mined. To evaluate Rule Growth and TRuleGrowth, we performed several experiments on four real-life datasets having different characteristics. First, the performance of Rule Growth was compared with CMRules and CMDeo while varying the *minsup* and *minconf* parameters, to as- sess their influence on the performance of each algorithm. Second, Rule Growth was compared to TRuleGrowth for different *window_size* values to evaluate the benefits of using the window size constraint. Experimental results show that RuleGrowth is up to several of magnitude faster and uses

up to an order of magnitude less memory than CMRules and CMDeo. Moreover experiment show that the execution time and the number of valid rules found can be reduced by several orders of magnitude when the window size constraint is used. Lastly, we have reported results from a real application where using par-tially-ordered sequential rules and the window constraint greatly improves accuracy over sequential rules.

REFERENCES

[1] U. Faghihi, P. Fournier-Viger and R. Nkambou, "A Computational Model for Causal Learning in Cognitive Agents," *Knowledge Based Sys- tems*, vol. 30, pp. 48-56, 2012.

[2] P. Fournier-Viger, U. Faghihi, R. Nkambou and E. Mephu Nguifo, "CMRules: An Efficient Algorithm for Mining Sequential Rules Com- mon to Several Sequences," *Knowledge Based Systems*, vol. 25, no. 1, pp. 63-76, 2012.

[3] P. Fournier-Viger, Knowledge discovery in problem-solving learning activities, Ph.D. Thesis, Univ. Quebec in Montreal, Montreal, 2010.

[4] P. Fournier-Viger, A. Gomariz, M. Campos and R. Thomas, "Fast Vertical Sequential Pattern Mining Using Co-occurrence Information," *Proc. 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, pp. 40-52, 2014.

[5] T. Bogon, I.J. Timm, A.D. Lattner, D. Paraske vopoulos, U. Jessen, M. Schmitz, S. Wenzel, S. Spieckermann, "Towards Assisted Input and Output Data Analysis in Manufacturing Simulation: The EDASIM Ap- proach," *Proc. 2012 Winter Simulation Conference*, pp. 257-269,

[6] M.A.Sartor,V.Mahavisno,V.G.Kesha mouni, J. Cavalcoli et al., "ConceptGen: a geneset enrichment and gene set relation mapping tool," *Bioinformatics*, vol. 26, no. 4, pp. 456-463, 2010.

[7] D. Lo, G. Ramalingam, V. P. Ranganath and K. Vaswani, "Mining Quantified Temporal Rules: Formalism, Algorithms, and Evaluation," *Proc. 16th Working Conference on Reverse Engineering*, pp. 62-71, 2009.

[8] P. Fournier-Viger, T. Gueniche and V.S. Tseng, "Using Partially- Ordered Sequential Rules to Generate More Accurate Sequence Predic- tion," *Proc. 8th International Conference on Advanced Data Mining and Ap- plications*, Springer, pp. 431-442, 2012.

[9] B. Kamsu-Foguem, F. Rigal and F. Mauget, "Mining association rules for the quality improvement of the production process," *Expert Systems and Applications*, vol. 40, no. 4, pp. 1034-1045, 2013.

[10] M.Nayrolles,N.Moha,P.Valtchev,"Improving SOA antipatterns detection in Service Based Systems by mining execution traces," *Proc.20th IEEE Working Conference on Reverse Engineering*, pp321-330, 2013.

Weighted Itemset Mining from Bigdata using Hadoop

Divya.M.G[#], Nandini.K[#], Priyanka.K.T[#], Vandana.B*
[#]*UG Student, *Assistant Professor*
Department of CS&E, RRCE, Visvesvaraya Technological University, INDIA
Divyashobha11@gmail.com, k.nandini95@gmail.com, priyankakt07@gmail.com, vandanacse@yahoo.co.in

***Abstract*—** Data items have been extracted using an empirical data mining technique called frequent itemset mining. In majority of theapplication contexts items are enriched with weights. Pushing an item weights into the itemset extraction process, i.e., mining weighted itemsets rather than traditional itemsets, is an appealing research direction. Although many efficient weighteditemset mining algorithms are available in literature, there isa lack of parallel and distributed solutions which are able to scale towards Big Weighted Data. This Proposed work presents an efficient frequent weighted itemset mining algorithm based on the MapReduce paradigm. It adopts the MapReduce architecture to partition thewhole mining tasks into smaller independent subtasks and uses Hadoop distributed file system to manage distributed data so that it allows the parallel and distributed solution.To demonstrate its actionability and scalability, the proposed algorithm will be tested on a Bigdataset collecting large volume of reviews ofitems. Weights indicate theratings given by users to the purchased items. The mined itemsets represent combinations of items that were frequently bought together with an overall rating above average.

Keywords-*MapReduce, Parallel Computing, hadoop, frequentitemset, Data mining, Distributed Computing, Apriori Algorithm.*

1. INTRODUCTION

The swift growth of data generated and stored has directed us to the new era of Big Data. Nowadays, we are surrounded bydifferent kinds of big data, such ase-commerce platform, sensor data, machine-generated data and social data. Extracting valuable information and insightful knowledge from big data has become an urgent need in many disciplines. In view of this, *big data analytics* [3, 4,] has emerged as a novel topic in recent years. This technology is particularly important to enterprisesand business organizations because it can help them to increase revenues, retain customers and make more intelligent decisions. Due to its high impact in many areas, more and more systems and analytical tools have been developed for big data analytics, such as *Apache Mahout*. And MLlib [5] are notable examples of MapReduce- and Sparkbased scalable machinelearning.

In real-life applications items are unlikely to be equally important within the analyzed data. For example, items purchased at the market have different prices, medical treatments have different urgency levels, and genes are expressed in biological samples with different levels of significance. Hence, an appealing extension of traditional itemset mining algorithms is to push of item relevance weights into the mining process. We propose Parallel Weighted Itemset miner, a new parallel and distributed framework to extract frequent weighted itemsetsfrom potentially very large transactional datasets enriched with item weights. The framework relies on a parallel and distributed-based implementation running on a Hadoop cluster. To make the mining process scalable towards Big Data, most analytical steps performed by the system are mapped to the MapReduce programming paradigm.

The Changes in the data access patterns of applications and the need to scale out to thousands of commodity machines led to the birth of a new class of systems referred to as weighted stores [8, 10, 18] which are now being widely adopted by

various enterprises. In the domain of data analysis, the MapReduce paradigm [17] and its open-source implementation Hadoop [20] has also seen widespread adoption in industry and academia alike. Solutions have also been proposed to improve Hadoop based systems in terms of usability [1, 28] and performance.

2. RELATED WORK

Different papers describes concept of data set mining as follows: (i) Large-scale item set mining (ii) weighted itemset mining. In large scale item set mining, exploratory Frequent itemset and association rule mining are widely data mining techniques which are first introduced in [6]. To scale towards large datasets most significant efforts have been devoted to studying parallel and distributed itemset mining strategies. For example, an Apriori-based [9] approach to mining frequent item sets has been presented in. Since Apriori is known to be less scalable than projection-based and vertical algorithms on complex datasets [10], many attempts to parallelize and distribute different itemset mining strategies have also been made (e.g., [12], [13]). For example, BigFIM is a hybrid algorithm based on MapReduce, which combines principles from both Apriori [9] and Eclat [11]. Solutions relying on FP-Growth like strategies have also been proposed. For example, the authors exploited prefix-tree-like structures to drive theparallel itemset mining process. The mining process entails the following steps: first, a horizontal subset of the data is analyzed and a local FP-Tree is built; then the item set mining process is performed on the local FP-Tree. Finally, the candidate pattern bases from different processing flows are then merged together. In an enhanced strategy for merging

processing flows has been proposed, more recently, the Parallel FP-Growth algorithm [13] parallelizes different instances of the recursive FP-Growth process on distributed machines. The key idea is to partition the computation in such a way that each machine executes an independent group of mining tasks thus reducing the communication between machines. To balance the computation load on different machines the authors proposed to consider the support of singletons, while the works presented and exploited clustering techniques and data sampling to limit the computational complexity of each mining task. An attempt to discover misleading patterns from Big datasets using MapReduce has been made in [14]. The idea is to compare frequent (unweighted) item sets mined at different abstraction levels to highlight potentially critical situations. Unlike all the aforesaid approaches, this paper addresses weighted itemset mining instead of traditional itemset mining. To scale towards towards Big Data, the Proposed framework relies parallel and distributed-based implementation running on a Hadoop cluster, where most mining steps are mapped to the MapReduce programming paradigm. Weighted itemset mining. In the traditional itemset and rule mining tasks items belonging to each transaction of the source dataset are treated equally. To differentiate items based on their relevance within each transaction, in [15] the authors first addressed the issue of mining more informative association rules, i.e., the Weighted Association Rules (WAR). WARs are association rules enriched with weights denoting item significance. Weights were introduced only during the rule generation step after performing the traditional frequent itemset mining process. To improve the efficiency of the mining process, the authors in [16] pushed item weights deep into the itemset mining process by exploiting the anti-monotonicity of the weighted support measure in an Apriori-based itemset mining process [9]. In [18] a FP-Growth-like weighted itemset mining algorithm process is presented.Unlike [15], [16], the algorithm proposed in [18] extracts infrequent (rare) item sets rather than frequent ones. A parallel issue is the extraction of weighted item sets and rules when coping with data not equipped with preassigned weights. For example, to generate appropriate item weights, in [17] the dataset is modeled as a bipartite hub-authority graph and evaluated by means of a well-known indexing strategy.

3. SYSTEM DESIGN

3.1.Hadoop Framework

Hadoop framework is allows data storing and running applications on clusters of commodity hardware [28]. It provides massive storage for any kind of data. Hadoop is the parallel programming platform built on Hadoop DISTRIBUTED File Systems (HDFS) for MapReduce computation. HDFS is highly fault-tolerant and is designed to bedeployed on low-cost hardware. HDFS holds very large volume of data and provide easier access. HDFS also makes applications available to parallel processing. HDFS is a part of Apache Hadoop main project.

3.2.MapReduce

MapReduce is a programming model and an associatedimplementation for processing and generating large datasets.Users specify a *map* and *reduce* functions, they takes <*key*,*value*>pair as an input and generates intermediate <*key*,*value*>pairs and merges all intermediate values associatedwith the same intermediate key respectively.

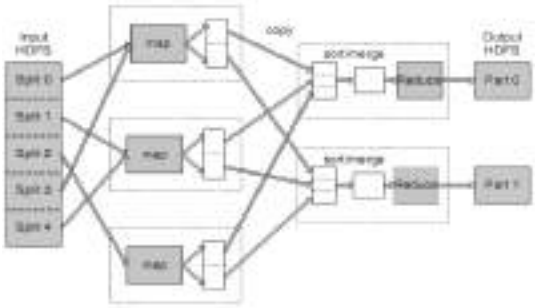


Fig. 1 Hadoop MapReduce Architecture.

3.3.Weigted ItemSet Mining System Architecture

This paper presents a scalable frequent weighted itemset mining algorithm based on the MapReduce paradigm. To demonstrate its action ability and scalability, the proposed algorithm will be tested on a Bigdataset collecting large volume of reviews of items. Weights indicate the ratings given by users to the purchased items. The mined itemsets represent combinations of items that were frequently bought together with an overall rating above average.

It integrates a variant of the BigFIM algorithm which is able to successfully cope with data enriched with weights. Furthermore, to allow experts to effectively explore the result of the mining process, the proposed system allows us to rank itemsets by (i) weighted support, (ii) traditional support, and (iii) a mix of the above. While the traditional support indicates the generic degree of interest of the considered combination of items, the weighted support integrated in the proposed framework indicates the average level of interest of the least interesting item within each transaction. The proposed system, running on an Hadoop cluster, overcomes the limitations of state-of-theart approaches in coping with datasets enriched with item weights.

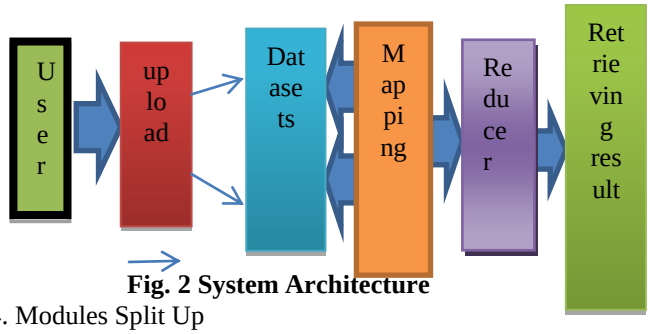


Fig. 2 System Architecture

3.4. Modules Split Up

Module 1: Registration module

This module is used for the user to register their login id by providing the minimal information. so that they can login to the website.
Module 2: Sign in module
In this, user can login to the website by registered login id and a valid password. Only the authenticated user can login and use the website.
Module 3: User module
In this module, user extracts itemset which are frequently brought together with an overall rating above average.
Module 4: Admin module
In this module, admin checks the item list, add the items and remove the unwanted items.
Module 5: Insert Dataset module
The large numbers of itemsets will be inserted to the database.

4. PARALLEL WEIGHTED ITEMSET MINING FROM BIGDATA

Parallel Weighted Itemset Miner is a new data mining environment aimed to analyze Big Data equipped with item weights. The main environment blocks are briefly introduced below. A more detailed description is given in the following section.

4.1. Data preparation

This step entails preparing data to the subsequent item set mining process. The source data is acquired, stored in a transactional dataset, and equipped with item weights. A transactional datasetis a set of transactions. Each transaction is a set of (not repeated) tems. Depending on the context of analysis, items may represent different concepts (e.g., products, objects, places, stocks). For example, let us consider the dataset reported in Table I. It is an example of (unweighted) transactional dataset consisting of five transactions, each one representing a different customer ofa e-commerce company. For each customer the list of purchased items is known. For instance, customer with id 1 bought items *X*, *Y*, and *Z*. Note that each transaction, which represents a distinct electronic basket, may contain an arbitrary number of items. To consider the relative importance of the items within each transaction during the item set mining process, itemsare enriched with weights. A transactional dataset whose items are equipped with weights is denoted as weightedtransactional dataset.

TABLE I: EXAMPLE OF UNWEIGHTED DATASET: ITEM BOUGHT BY CUSTOMERS

Customer id	Purchased items
1	*X*, *Y* , *Z*
2	*X*, *Y* , *Q*
3	*X*, *Y* , *Z*
4	*X*, *Y* , *Q*
5	*X*, *W*, *Z*

TABLE II: EXAMPLE OF WEIGHTED DATASET: ITEM RATINGS GIVEN BY CUSTOMERS

Customer id	Purchased items and ratings
1	*(X*, 3*)(Y*, 1*) (Z*, 5*)*
2	*(X*, 2*)(Y*, 2*) (Q*, 2*)*
3	*(X*, 4*) (Y*, 2*) (Z*, 5*)*
4	*(X*, 3*) (Y*, 3*) (Q*, 2*)*
5	*(X*, 2*)(W*, 5*)(Z*, 4*)*

item is a pair *_item*, *weight_*, where *weight* is the weight associated with the corresponding *item*. For example, let us consider the weighted transactional dataset reported in Table II. It extends the traditional transactional dataset in Table I by enriching items with the corresponding weights. More specifically, for each customer the rating (from one to five) given to each purchased item is known. For instance customer with id 1 rated item *X* as 3, item *Y* as 1, and item *Z* as 5. The analyzed data are tailored to a weighted transactional data format. Furthermore, if need be, ad hoc preprocessing steps are applied to the raw data to ensure high-quality results. For example, data filtering and discretization areexamples of commonly used preprocessing steps [19]. Data filtering entails discarding the items/transactions that are irrelevant for subsequent analyses. For instance, recalling the previous example, duplicate entries of the same customer basket can be removed because they could bias item set support counts. To ensure the scalability of the knowledge discovery process, the PaWI system performs data filtering as a distributed MapReduce job.

4.2. Weighted Item setMining

This step focuses on mining frequent weighted itemsets[15] from the prepared weighted dataset. A *k*-itemset(i.e., an itemset of length *k*) is a set of *k* items. Thetraditional support value of an itemset in a transactional dataset is given by its frequency of occurrence in the source dataset [6]. For example, *{X,Y}* is an itemset indicating the co-occurrence of items *X* and *Y* . If we disregard item weights, this itemset has a support equal to 4 in Table I because it occurs in four out of five transactions, meaning that most of the users purchased items *X* and *Y* together. The goal of this paper is to extend traditional large-scale itemset miners to successfully cope with Big weighted data. Hence, for our purposes, the itemset support measure is extended, similar to [15], to the case of weighted data. As previously done in [18], the weighted support of an itemset*I* in a weighted transactional dataset *T* is defined as a linear combination of the aggregation weights computed on each transaction in *T*. An arbitrary aggregation function*f*(e.g., min, max, average, and mode) can be potentially applied to aggregate item weights within each transaction. The choice of *f* depends on the considered use cases. Hereafter, similar to [18], we will consider *f*=min (i.e., the least weight of any item in *I* is considered), because, as

discussed, the selected patterns are deemed as particularly useful for analyzing real Big datasets. Recalling the running example, let us consider analystswho would like to discover the combinations of items that were frequently bought together with an overall rating above average. To this aim, we may consider item ratings during support computation by weighting item set occurrences within each transaction by the least item rating. For instance, recalling the weighted transactional datasetin table for customer with id 1 between *X* and *Y* the item with least rating is *Y* (rating equal to 1), while for customer with id 5 is *X* (rating equal to 2). Hereafter we will denote as *weighted support* the support of an item set by considering item weights, whereas as *traditional support* the item set support disregarding item weights. For instance, *{X,Y}* has weighted support equal to 1+2+2+3+0=8 and traditional support equal to 4. Given a weighted transactional dataset *D* and an (analyst provided)minimum support threshold *minus*, the PaWIsystem addresses the extraction of all frequent weighted item sets from *D*. To allow comparing weighted item sets with traditional ones, PaWI allows experts to mine traditional item sets as well. As discussedbelow, the support thresholds enforced during weighted and unweighted item set mining are potentially different. The weighted item set mining process relies on a parallel and distributed-based algorithm running on a Hadoop cluster [20]. To make the mining process scalable towards Big Data, the mining steps are mapped to the MapReduceprogramming paradigm. MapReduce [21] is a parallel programming framework providing both a relatively simple programming interface together and a robust computational architecture. MapReduce programs consist of two main steps. In the map step, each mapper processes a distinct chunk of the data and produces key-value pairs. In the reduce step, key-value pairs from different mappers are combined by the framework and fed to reducers as pairs of key and value lists. Reducers generate the final output by processing the key/value lists. To efficiently perform frequent weighted item set mining with MapReducePaWI integrates a variant of the BigFIM algorithm [12] which is able to successfully cope with data enriched with weights. The exploitation of weights is challenging because ad-hoc data structures must be used to efficiently maintain the weights associated with each item and transaction by balancing the impacts on computational and communication costs. The following extensions have been proposed: Distributed transaction splitting. BigFIM relies on two established item set miners: Apriori [9] and Eclat [11]. We modified the BigFIM algorithm to allow both Apriori andEclat to successfully cope with weighted data. More specifically, our algorithm generates an equivalent version of the source dataset that includes only transactions with equally weighted items. Let us assume that the weight of an equivalent transaction *tq*is*w*. Then, the occurrence of any item set in *tq*will be weighted by *w* instead of by 1. Each transaction in the original dataset may correspond to a set of equivalent transactions in the equivalent dataset. A formal definition of the equivalence set of weighted transactions is given in [18]. Note that since two distinct transactions have disjoint equivalent sets the splitting process is straightforwardly parallelizable. Weighted support counting.Since items are equipped with weights, traditional support counting is replaced with weighted support counting, according to Definition 1. Toaccomplishitemsetsupport counting different strategies areadopted according to the algorithm used. Specifically, to perform Apriori-based weighted itemset mining, item set supports are counted by generalizing the word counting problem for documents [21] to weighted item sets, i.e., each mapper receives a disjoint subset of dataset transactions (i.e., the documents) and reports the items/item sets (i.e., the words) for which the weighted support count is performed. A reducer combines all partial weighted support counts and reports only the items/item sets whose weighted support is above the threshold. These frequent weighted item sets are redistributed to all mappers to act as candidates for the next step of breadth-first search [9] and then the procedure is repeated to mine weighted item sets of higher length. To perform Eclat-based weighted itemset mining, each mapper builds the weighted tidlist of the item sets related to a subset of transactions. The weighted tidlist of an arbitrary item *i j* consists of all pairs (*transaction id*, *weight*) such that the transaction related to *transaction id* contains item *i j* with weight. For example, let us assume that a mapper receives the transactions contained in the dataset in Table II. For item *Z* it generates the following weighted tidlist: *{*(cid1,5),(cid3,5),(cid5,4)*}*. The weighted tidlist consists of all pairs (*customer id*, *weight*) for which the transaction related to *customer id* contain item *Z* with weight. For instance, the transaction corresponding the electronic basket of the customer with id 1 contains item *Z* with weight 5. A reducer combines all partial weighted support counts and reports only the items/item sets whose weighted support is above the threshold. Note that the equivalent transaction weights are not stored in the distributed cache, because Bigdatasets may potentially consist of millions of transactions.

4.3. Item set Ranking

The manual exploration of all the item sets (weighted or not) mined from Big data is practically unfeasible. Hence, to support the knowledge discovery process experts may wouldlike to access only a subset of most interesting patterns. This step focuses on ranking the mined item sets according to their level of significance in the analyzed data. To filter and rank the mined item sets, the support measure is the most commonly used quality index [6]. To cope with weighted data, for each candidate itemset the PaWIsystem computes both the traditional and weighted support measures. While the traditional support value indicates the observed frequency of occurrence of the considered combination of items in the source dataset, in weighted support counting itemset occurrences are weighted by the least item weight (see Definition 1). To select item sets whose average least item weight is maximal the PaWI system combines the weighted and traditional support measure in a new measure called AW-

sup, i.e., the *Average Weighted support*. The AW-sup measure is defined as the ratio between the weighted emset support and the traditional itemset support. It indicates the average per-transaction weight of the least weighted item. Selecting top interesting item sets based on this measure is potentially
interesting in real applications. For example, let us consider again the example dataset in Table II. According to Definition 1, itemset*{X,Y}* has weighted and traditional support values equal to 8 and 4, respectively. Since transactions represent electronic baskets, the weighted itemset support indicates the overall least item rating computed on the subset of customers who bought both items *X* and *Y* , while the traditional support measure indicates the simple frequency of occurrence of the combinations of items in the electronic basket dataset. The AW-sup value of *{X,Y}* is 2, meaning that, on average, for each electronic basket containing items *X* and *Y* both items have been rated at least 2. Ranking the mined item sets by decreasing AW-sup allows experts to consider first the combinations of items that got maximal average ratings. Note that this statistics cannot be straightforwardly computed based on simple averages, because (i) it considers only the electronic baskets containing both *X* and *Y* , (ii) for each basket it selects the rating of the least weighted item between *X* and *Y* . Item sets not satisfying the traditional support threshold are discarded because they represent combinations of items that rarely occur in the analyzed data. The setting of the minimum weighted support threshold is driven by the average rating of the selected items. More specifically, we are interested in exploring the frequent combinations of items with rating above average, i.e., the item sets whose AW-sup is above a minimum threshold.

5. IMPLEMENTATION

The below steps gives an overview about weighted item set mining.

Algorithm

Step1: Collect the data sets
Step 2: Select any data from datasets
Step3: Stores datasets with weight (Rank or review)
Step4: Mapping part introduce in mining
Step5: Search parallel items in datasets.
Step6: In reducing part we can get result.

Fig. 3 Sign in module

Fig. 4 Sign up module

Fig. 5 Cloud module

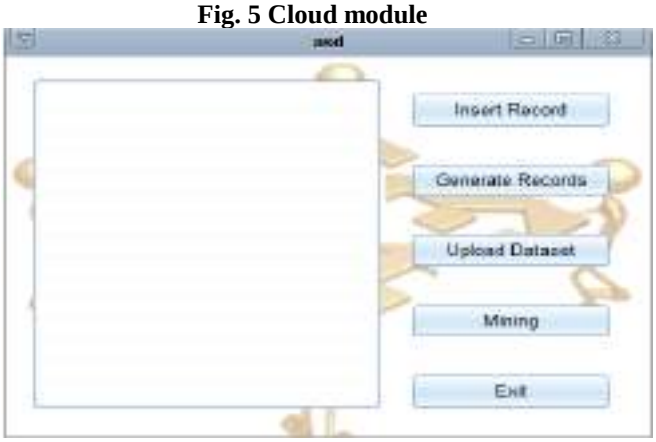


Fig. 6 User module

CONCLUSIONS AND FUTURE WORK

This paper presents a parallel and distributed solution to the problem of extracting frequent itemsets from Big Weighted Datasets. The proposed system, running on a Hadoop cluster, overcomes the limitations of state-of-theart approaches in coping with datasets enriched with item weights. The experiments, performed on a large volume of dataset, confirm the actionability of the mining result. Future works will entail the application of the proposed approach to recommender systems. For example, discovering combinations of items that were frequently bought together with an overall rating above average couldbeusefulfor recommending additional items beyond those already purchased by a given user.

REFERENCES

[1] D. Agrawal, S. Das, and A. El Abbadi, "Big data and cloud computing: Current state and future opportunities," in *Proceedings of the 14th International Conference onExtending Database Technology*, ser. EDBT/ICDT '11. New York, NY, USA: ACM, 2011, pp. 530–533.

[Online].Available:http://doi.acm.org/10.1145/1951365.1951432
[2] H. T. Lin and V. Honavar, "Learning classifiers from chains of multiple interlinked RDF data stores," in *IEEE InternationalCongress on Big Data, BigData Congress 2013, June 272013-July 2, 2013*, 2013, pp. 94–101. [Online]. Available: http://dx.doi.org/10.1109/BigData.Congress.2013.22
[3] J. Hartog, R. Delvalle, M. Govindaraju, and M. J. Lewis, "Configuring a mapreduce framework for performance heterogeneous clusters," in *2014 IEEE InternationalCongress on Big Data, Anchorage, AK, USA, June 27- July 2, 2014*, 2014, pp. 120–127.[Online].Available:http://dx.doi.org/10.1109/BigData.Congress.2014.2631
[4] "The Apache Mahout Project: The Apache Mahout machine learning library. Available: http://mahout.apache.org/ Last access: March 2015," 2015.
[5] "MLlib: Apache Spark's scalable machine learning library. Available: http://spark.apache.org/mllib/ Last access: March 2015.
[6] R. Agrawal, T. Imielinski, and Swami, "Mining association rules between sets of items in large databases," in *ACMSIGMOD 1993*, 1993, pp. 207–216.
[7] L. Cagliero and P. Garza, "Itemset generalization with cardinality-based constraints," *Information Sciences*, vol. 244, pp. 161–174, 2013.
[8] E. Baralis, L. Cagliero, S. Jabeen, and A. Fiori, "Multidocument summarization exploiting frequent itemsets," in *Proceedings of the ACM Symposium on AppliedComputing, SAC 2012, Riva, Trento, Italy, March 26-30, 2012*, 2012, pp. 782–786. [Online].Available: http://doi.acm.org/10.1145/2245276.2245427
[9] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," *in VLDB'94, Proceedingsof 20th International Conference on Very Large DataBases*, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds. Morgan Kaufmann, 1994, pp. 487–499.
[10] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACMSIGMOD International Conference on Management of Data*, 2000, pp. 1–12.
[11] J. Neville, D. Jensen, and B. Gallagher, "Simple estimators for relational bayesian classifiers," in Proceedings of the Third IEEE International Conference on Data Mining, 2003, pp. 609–612.
[12] C. Perlich and F. Provost, "Distribution-based aggregation for relational learning with identifier attributes," Machine Learning, vol. 62, no. 1-2, pp. 65–105, Feb. 2006.
[13] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," in AAAI-98 Workshop on Learning for Text Categorization. AAAI Press, 1998, pp. 41–48.
[14] T. Ferguson, "A bayesian analysis of some nonparametric problems," The annals of statistics, pp. 209–230, 1973.
[15] T. P. Minka, "Estimating a dirichlet distribution," Tech. Rep., 2012.
[16] R. E. Madsen, D. Kauchak, and C. Elkan, "Modeling word burstiness using the dirichlet distribution," in Proceedings of the 22nd International Conference on Machine Learning, ser. ICML '05. New York, NY, USA: ACM, 2005, pp. 545–552.
[17] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," Journal of Theoretical Biology, vol. 29, no. 3, pp. 471–481, 1970.
[18] M. Nickel, V. Tresp, and H.-P.Kriegel, "Factorizing YAGO: scalable machine learning for linked data," in Proceedings of the 21st international conference on World Wide Web, ser. WWW '12. New York, NY, USA: ACM, 2012, pp. 271–280.
[19] U. Losch, S. Bloehdorn, and A. Rettinger, "Graph kernels for ¨ RDF data," in The Semantic Web: Research and Applications, ser. Lecture Notes in Computer Science, E. Simperl, P. Cimiano, A. Polleres, O. Corcho, and V. Presutti, Eds. Springer Berlin Heidelberg, 2012, vol. 7295, pp. 134–148.

Live Website Traffic Analysis Integrated with Improved Performance for Small Files using Hadoop

Navyashree B L[#1], Rakshitha H[#2], Ranjitha M[#3], T Auntin Jose [#4]
1,2,3 Bachelor of Engineering Students, 4 Associate Professor
Department Of Computer Science And Engineering,
RAJARAJESWARI COLLEGE OF ENGINEERING
BENGALURU-560074, KARNATAKA, INDIA
navyashree333@gmail.com, rakshitha.h@gmail.com, ranjitha200008@gmail.com, auntin123@yahoo.com

Abstract— Hadoop, an open source java framework deals with big data. It has HDFS (Hadoop distributed file system) and MapReduce. HDFS is designed to handle large amount files through clusters and suffers performance penalty while dealing with large number of small files. These large numbers of small files pose a heavy burden on the NameNode of HDFS and an increase execution time for MapReduce. Secondly, as an application part Traffic analyzer implemented with the combination of Hadoop and Map-Reduce paradigm, which makes it possible to analyse the any website programmatically. A web ranking metric, web analytics or simply web measurement refers to a system used to measure factors that affect a website's exposure and traffic on the web. The proposed approach is done to handle small files. In proposed approach, "merging" of small file is done using MapReduce programming model on Hadoop. This approach improves the performance of Hadoop in handling of small files and also reduces the memory required by NameNode to store them. Traffic analysis gives the rank, number of views, visitors, index number so on any website which indicates the true analysis of the website in frequent basis using Hadoop.

Keywords— *MapReduce; Hadoop; HDFS; Small Files; Traffic Analyzer;*

I. INTRODUCTION

A.Hadoop

As Internet is growing rapidly, data is exploding from and growing beyond limit of expectation, the traditional techniques like RDBMS miserably fail to handle large files. Apache Hadoop [1][2] was created out of necessity by Doug Cutting. Published in the papers by Google, Hadoop was inspired in storing, processing and analyzing hundreds of terabytes, and even petabytes of data [3]. No data is too big for Hadoop.

The architecture of Hadoop is as shown in the figure1.

i. HDFS

a. Name Node

Apache Hadoop includes HDFS (Hadoop Distributed File System) is distributed file system designed .This enables the data to be processed in parallel using all of the machines in the cluster. HDFS is a portable file system written in Java for the Hadoop framework.

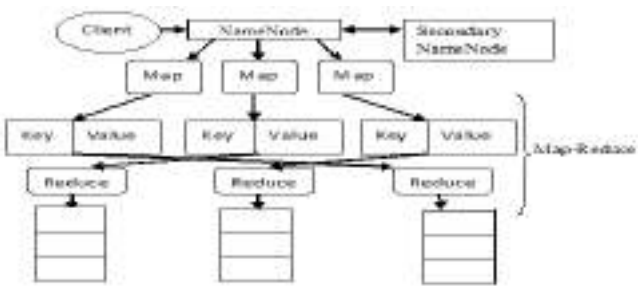


Figure1: Basic architecture of Hadoop

The design of the HDFS is such that it is used to store large files. The architecture of HDFS design stores large cluster of data in form of NameNode and DataNode which is taken from client. It consists of a metadata in the form of the NameNode and a large number of I/O nodes called DataNode.

b. Data Node

The DataNode works as the slave on which the actual data resides. To indicate its presence in the system, the DataNode keeps on sending status signal to the NameNode at regular intervals. The DataNode is responsible for serving the read and write request for the client. The daemon named as TaskTracker runs on the DataNode which is responsible for executing the individual tasks assigned by the JobTracker.

The DataNode service all read/write and file replication requests based on job given from the NameNode. Because Hadoop keeps all file system metadata in main memory, it is necessary for the NameNode to be its own server, this way file access is not slowed because of strain on the NameNode from serving metadata requests.
In order to keep the replication high and to rebalance the data, the DataNodes interact with one another and moves and copies the data around.

c. HDFS Client

Client machines have Hadoop installed with all the settings, but are neither a Master nor a Slave. Client machine loads the data into the Hadoop cluster, along with this it also invokes the map reduce , which says the

processing of data and also view or retrieved the results when the process is done. Client can read, write and delete files and also perform the operations to create and delete directories with contacting to NameNode. Based on the Client request only the NameNode process the job.

ii . Map Reduce

MapReduce is another component of Hadoop. MapReduce is regarded as the heart of Hadoop. In MapReduce, there are two job that Hadoop performs i.e. map() function and reduce() function. The map() function is designed to take the set of data and breaks into number of different tuples in the form of key-value pair. Since there are multiple DataNodes where the huge data is stored, processing the required file among huge set of file is a time consuming task, as solution to this the key-value pair is generated where searching of data becomes easier.

The reduce() function on the other hand considers the output from the map as an input and then combines those data tuples into a smaller set of tuples. The map job is always performed prior to the reduce job.

B.Traffic Analyzer

A website tells a lot about a business. It shows how much thought the business puts into its brand and whether it values having a website. Unfortunately, far too many companies don't really value their websites and don't get the full benefit out of them. They neglect design, website copy, and other important essentials. They put the focus only on making sales.

This results in a really bad website and leaves visitors unsure if the company is the best one to do business with. This is why it's so important to have a value proposition. A website needs to tell visitors in a couple of sentences or less why their business is the best choice for the visitor, instead of sending a bunch of different messages that won't be received.

Unlike other marketing venues, visitors to a website are typically anonymous. Web traffic analyzer helps to understand the behavioral data using Alexa server. The analyzer draws the data from the server calculates the result and generate the files which contains the calculated result.

II. PROBLEM ANALYSIS

Hadoop is a better approach is mainly designed to handle the huge amount of data in form of terabytes anta petabytes. But it has a major performance issue for handling with large number of small files. The data is stored using HDFS in multiple nodes, the default block size of HDFS is 64MB designed to access large files. If there is a file whose size is less than 64MB it is stored on one entire block. Consider the case which is shown in the diagram, where Client request for analysis of few set of websites. Traffic analyser pull the registered website from the database, it communicates through the Alexa server and fetches the information from server and generates the report in the form of a file for given websites. These files might be of size 1KB. If there are several files generates of these size is put on to the Hadoop cluster. These files are processed and each file is stored different blocks, where the remaining memory is being unused. This results in inefficiency of memory management.

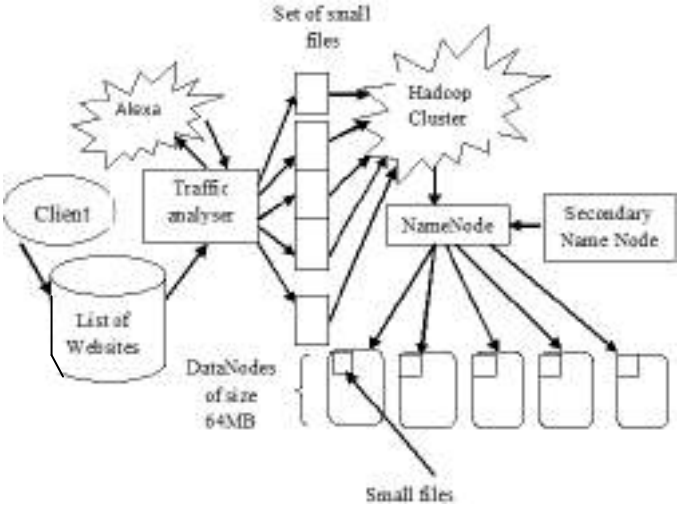


Figure 2: Problem in HDFS

Along with the memory being inefficient, small files pose a heavy burden on the NameNode. To read or access the data from a small file causes hopping from DataNode to DataNode and a lot many seeks are required in search of data from the small file. In case there are large number of small files, then each map task processes very little input and for every data there is a map tasks.

It imposes a high overhead in the system. In the NameNode, capacity of namespace of the system is limited by physical memory. This causes metadata to occupy larger portion of the memory.

So, if there are thousands of small files than the efficiency of HDFS fail to manage the files via performance, memory management becomes disaster.

III. PROPOSED APPROACH

According to the problem analysis, Hadoop suffers from performance issues from large number of small files. To overcome this problem, "Merging" solution is taken. But merging algorithm is time consuming. In order to reduce the time consumption, small files can be combined parallel using the MapReduce algorithm. In map-Reduce algorithm threshold value can be set, whose value will be slightly lesser than block size of HDFS. In this algorithm, Map() function will fetch the file, generates the key-value pair, where the key is file size and value is filename . This key-value pair will be given to Reduce() function. The Map() would keep on adding the files until the default block size is reached and then pass it to Reduce(). The Reducer will merge the files. This process will then be carried out parallel till all the files are combined to the default block size. Since it is carried parallel, this reduces

the time for merging and executing the files. If the file is larger than threshold value it ignores those files.

Consider the case mentioned in the problem analysis, according to the solution mentioned the architecture diagram is shown in figure 2.

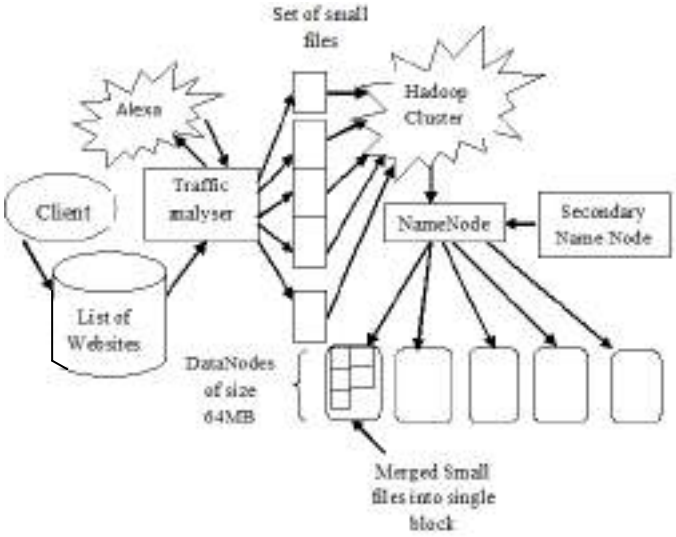


Figure 2: Problem in HDFS

In this approach, the files being sent to the NameNode generates the key-value pair. If the key i.e. the size is relatively lesser than the threshold being set in the algorithm, then the files are merged till it reaches the threshold value and kept in the DataNode. This clearly shows that the memory is efficiently used after processing. NameNode burden for execution also reduces. Traffic analyser is also proposed using Map-reduce algorithm this helps to have the minimum response time. The computing capacity of the website analysis also reduces.

Through this approach the execution time will reduce which is shown through graph in figure 3.

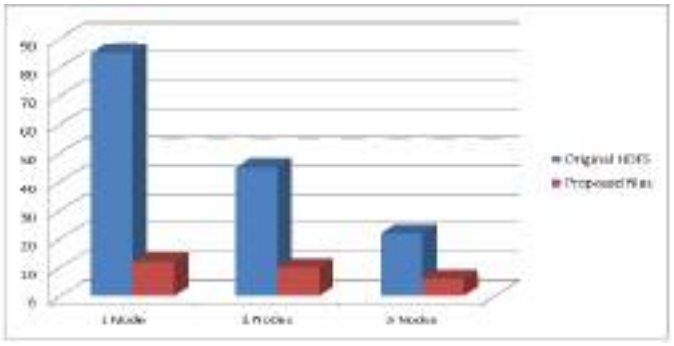


Figure 4: Comparison of Different Approaches of execution time on different number of nodes

Original HDFS takes a lot of time to execute the application for small files as it is the rule of thumb, Hadoop that each file is separately stored in individual block and each block is mapped to one map. So an increase in the number of map() function, increases the execution time. It is depicted in figure 4.

Consider the case to having 20000 small files. The size of these files range from 10KB to 126MB. The cumulative size of all the files is approximately 525MB.

According to proposed approach they are treated as individual file and store in its metadata into main memory. So it requires very less amount of memory than original HDFS. It is depicted in figure 5.[3]

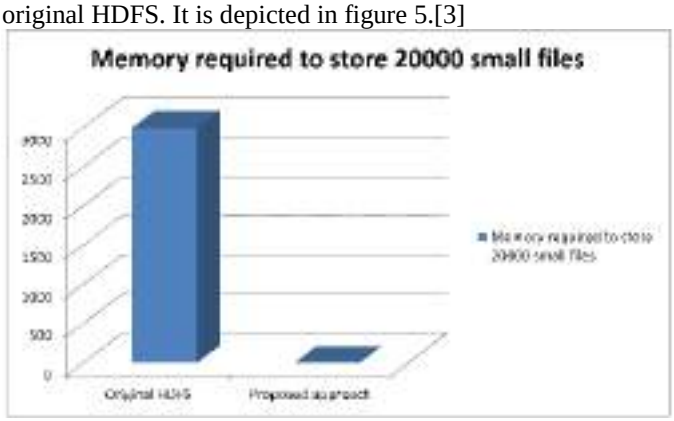


Figure 5: Memory required for storing small files

IV. CONCLUSION AND FUTURE WORK

As such Hadoop being wide area of research and one of the topics is chosen for research is handling of small files in HDFS, so the following research focuses on a MapReduce approach to handle small files, considering mainly two parameters. Firstly, execution time to run file on Hadoop Cluster and secondly the memory utilization by NameNode. By considering these parameters, proposed algorithm improves the result compared to existing approaches. Also the analysis of website traffic using Map-Reduce paradigm reduces the computing time and helps for the parallel processing. Thus, the overall system helps to increases the efficiency of the Hadoop for generated small files.

As for future work, small file storage solutions on HDFS will be mainly studied for other types of files as well. Based on file type's analysis, small files are classified as multiple types, and customized approaches will be supplied to different types to further improve the efficiency. The Solution might also work on the merging of different types of files and obtaining the efficiency of the algorithm. Image files also can be considered because it is also a small file whose performance is low in the HDFS.

REFERENCES

[1] http://hadoop.apache.org/core/.

[2] Jeffrey Dean and Sanjay Ghemawat, MapReduce: Simplied Data Processing on Large Clusters, Google, I nC.2004

[3] Dhruba Borthakur, The Hadoop Distributed File System: Architecture and Design, ACM 2006

[4] Parth Gohil, Bakul Panchal, J S. Dhobi "A Novel Approach to Improve the Performance of Hadoop in Handling of Small Files" 2015 IEEE

[5] Grant Mackey, Saba Sehrish, Jun Wang "Improving Metadata Management for Small Files in HDFS" 2009 IEEE

[6] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler "The Hadoop Distributed File System" IEEE 2010

[7] JaiPrakesh Varma, Bamkim Patel, Atul Patel, "Big Data Analysis: Recommendation System with Hadoop Framework" , in 2015 IEEE International Conference on Computational Intelligence & Communication Technology

[8] Liu Jiang, Bing Li , Meina Song "THE OPTIMIZATION OF HDFS BASED ON SMALL" Proceedings of IC-BNMT 2010

[9] http://www.cloudera.com/blog/2009/02/02/the-small-files-problem.

[10] J. Venner, Pro Hadoop. Apress, June 22, 2009.

[11] Http://issues.apache.org/jira/browse/HADOOP-1687.

[12] Hadoop archives, http://hadoop.apache.org/common/docs/current/hadoop_archives.html.

Customer behavior analysis of web server logs using Hive in Hadoop Framework

Lavanya KS, Srinivasa R
Dept. of Studies in CSE, RRCE College of Engineering, Bangalore, Karnataka, India
Email: lavanyaa.ks@gmail.com , srinivasa.r@gmail.com

Abstract: Web log file is a log file created and stored by a web server automatically. Analyzing such web server access logs files will provide us various insights about website usage. Due to high usage of web, the log files are growing at much faster rate with increase in size. Processing this fast growing log files using relational database technology has been a challenging task these days. Therefore to analyze such large datasets we need a parallel processing system and a reliable data storage mechanism (Hadoop). Hadoop runs the big data where a massive quantity of information is processed via cluster of commodity hardware. In this paper we present the methodology used in pre-processing of high volume web log files, studying the statics of website and learning the user behavior using the architecture of Hadoop MapReduce framework, Hadoop Distributed File System, and HiveQL query language.
Keywords: big data, customer behavior analysis, hadoop, log analysis, web server logs

1. INTRODUCTION

In today's competitive environment, manufacturers/service providers are keen to know whether they provide the best service/product to customers or whether customers look forward to get their service or buy product. Service providers should need to know how to make their websites or web application interesting to customers and how to improve advertising strategies to attract them. All these questions can be answered by Log files. Log files contain a list of actions that occurred whenever customer accesses the service provider's website or web application. Every "hit" to the Website will be logged in a log file. These log files are stored in web servers. The raw web log file is one line of text for each hit to the website and contains information about who visited the site, where they came from, and what they did on the website. These log files carry a useful information for service providers so analyzing these log files can give them insights about website traffic patterns, user activity, customer interest etc.

2. RELATED WORK

As data center generates thousands of terabytes or petabytes of log files a day it is highly challenging to store
and analyze such high volumes of log files. Analyzing log files looks complicate because of their high volume and

complicate structure. Traditional database techniques have failed to handle these log files efficiently due to large size. In 2009, Andrew Pavlo and Erik Paulson compared the SQL DBMS with Hadoop MapReduce and
suggested that Hadoop MapReduce loads data sooner than RDBMS. Also traditional RDBMS cannot handle large datasets.

This is where big data technologies play a major role in handling large sets of data. Hadoop is the best suitable platform that stores log files and does parallel

implementation of MapReduce program. For enterprises Apache Hadoop is a new way to store and analyze data.

Hadoop is an open-source project created by Doug Cutting under the administration of the Apache Software Foundation. Hadoop enables applications to work with thousands of nodes with petabytes of data. While it can be used on a single machine, its true capability lies in scaling to hundreds or thousands of computers. Tom White describes Hadoop is specially designed to work on large volume of data using commodity hardware in parallel. Hadoop breaks log files into equal sized blocks and these blocks are evenly distributed among thousands of nodes in cluster. Further, it does the replication of these blocks over multiple nodes to provide reliability and fault tolerance.

In case of large log files parallel computation of MapReduce improves performance by breaking job into many tasks. Hadoop implementation shows that MapReduce program structure can be an effective solution to analyze large volume of weblog files in Hadoop environment. In my project Hadoop-MR log file analysis tool, which provides a statistical report on total hits of a web page, traffic sources and user activity, was performed on two machines with three instances of Hadoop by distributing log files evenly among all nodes. A generic log analyzer framework for different kinds of log files was executed as distributed query processing to minimize response time for the users that can be extendable for some format of logs. Hadoop framework handles large volume of data in a cluster for web log mining. Data cleaning, major part of preprocessing was performed to remove inconsistent data. The preprocessed data was again manipulated using session identification algorithm to explore the user session. Unique identification of fields was carried out to track the user behavior.

3. HADOOP MAP REDUCE

Hadoop is an open source framework used for large scale computation and processing on a cluster of commodity hardware. It permits applications to work along with thousands of independent computers. The important characteristic of Hadoop is to move computations on the data rather than move data for computation. Hadoop is mainly used to breakdown large amount of input data into smaller chunks and each can be later processed separately on different computers. It implements a MapReduce programming model to achieve parallel execution.

MapReduce is a java based distributed programming model consisting of two phases: a parallel "Map" phase, followed by an aggregating "Reduce" phase. Map function

processes a key/value pair (k1, v1, k2, v2) to create a set of intermediate key/value pairs. On the other hand, reduce function merges all intermediate values [v2] that are associated with the same intermediate key (k2).Map (k1, v1) → [(k2, v2)],Reduce (k2, [v2]) → [(k3, v3)].Maps are the individual tasks that convert input records into intermediate records. A MapReduce task usually splits the input data into individualistic chunks that are then processed by the map tasks. The framework sorts the output data of the map that are then sent to reduce tasks. Both the input and the output data are stored in the Hadoop file-system.

The Hadoop cluster carry a single NameNode, a master that manages the file system namespace and synchronize its access to files by clients. There can be number of DataNodes (usually one per node in the cluster) that report the list of blocks it stores to NameNode periodically. HDFS replicates files for a configured number of times and re-replicates automatically the data blocks on nodes that have failed. With the help of HDFS any file can be created, copied and deleted but cannot be updated. The file system uses TCP/IP to communicate between the clusters.

4. PROPOSED METHODOLOGY AND DISCUSSIONS

Log files normally generated from the web server consist of large volume of data that cannot be handled by a traditional database or additional programming languages for computation.

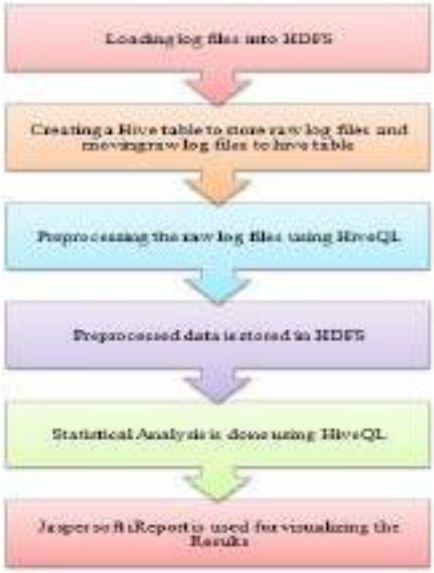


Fig.1 flow chart describing the methodology

The proposed work aims on preprocessing the log file using Hadoop is shown in Fig 1.The work is split into phases, where the storage and processing is made in HDFS.

Web server log files are first copied into Hadoop file system and then loaded to Hive table. Data cleaning, which is done using Hive query Language, is the first phase carried out in our project as a pre-processing step. Web server log files consist of number of records that correspond to automatic requests generated by web robots. The records usually carry a large volume of misleading, erroneous, and incomplete information. In our project web log files carrying requests from robots, spider and web crawlers are removed. Notably, requests created by web robots are not considered as useful data and are filtered out from the log data

In preprocessing the entries that have a status of "error" or "failure are removed. Further few access records generated by automatic search engine agent are identified and eliminated from the access log. The identification of status code is the important task carried out in the data cleaning .Only the log lines with the status code value of "200" are identified as correct log. Therefore only the lines with the status code value of "200" are extracted and stored in Hive table for further analysis.

The next step is to identify unique user, unique fields of date, status code, and URL referred. These unique values are retrieved and used for further analysis to find the total URL referred on a particular data or the maximum status code with successes on precise date.

In our project Hadoop framework is used to compute the log processing through pseudo distributed mode of cluster.The web server logs of www.ubdtce.org(for a period of five months from December 2014 to March 2015) are used for processing in Hadoop framework.the log files are mainly analyzed with the help of Centos 6.6 OS with Apache Hadoop 1.1.2 and Apache Hive 0.10.0.

4.1 Pseudo Distributed Mode.

Hadoop framework incorporate five daemons namely Namenode, Datanode, Jobtracker, Tasktracker, Secondary namenode. In case of pseudo distributed mode all the daemons are run on local machine that actually stimulates a cluster.

4.2 Apache Hive.

Apache Hive [13] is an important tool in the Hadoop ecosystem that provides a Structured Query Language called HiveQL to query the data stored in Hadoop Distributed File system. The log files that are stored in the HDFS are loaded in to a hive table and cleaning will be performed. The cleaned web log data is used to analyze daily statistics ,unique user and unique URLs, monthly statistics etc.

4.3 JasperSoftiReport Designer.

JasperSoftiReport Designer is a powerful graphical design tool used by report designers. iReport can help to design reports to meet a myriad of retorting needs. iReport is mainly built on the NetBeans platform and is available as a standalone application or as a Netbeans plug-in. After pre-processing, by making a JDBC connection to Hive jaspersoft'siReport 5.6 the results stored in HDFS is visualized in the form of graphs and tables.

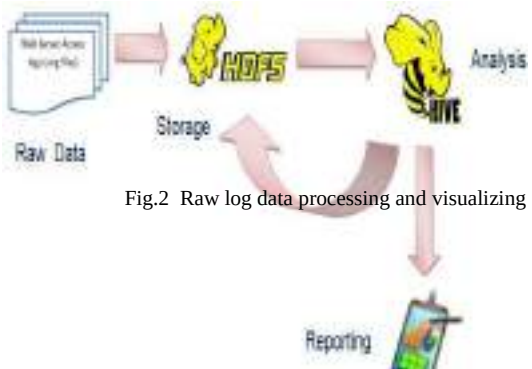


Fig.2 Raw log data processing and visualizing

The above flow chart illustrates copying raw log files in to HDFS and then preprocessing is done with the help of Apache Hive data warehouse tool. Next JasperSoft's iReport tool is used to produce the analysis results in the form of graphs and tables.

5.EXPERIMENTAL RESULTS

One of the main advantage of data cleaning is it produces quality results with increased efficiency. The results from Pre-processing step are shown in table below. The results indicate how much data reduction in size has taken place.

	Raw Data	After Cleaning
File Size	108.4 MB	9.3 MB
No. of Rows	4, 66,621	47, 039

Table 1. Results Before And After Pre-Processing

In our project web access logs were taken from www.ubdtce.org website for time period of 31/oct/2014 to 31/mar/2015 and the following results were obtained.

5.1 General Statistics

In this section we get general information related to the website like how many times the website was hit, total number of visitors, bandwidth used etc. It lists out all the information that one should know about the websites. The below table indicates number of hits, visits and bandwidth consumption of ubdtce.org website for a period of five months.

Summary	
Hits	
Total Hits	466621
Visitor Hits	422591
Visitors	
Total Visitors	47039
Total Unique IPs	4560
Bandwidth	
Total Bandwidth	8698.03 MB
Visitor Bandwidth	8219.00 MB

Table 2. General Statistics accessed After Analyzing Web Logs

5.2 Activity statistics

This section provides statistics on daily and monthly basis. It also gives on what days the website had visited maximum. Fig.2 and Fig.3 shows the daily and monthly obtain statistics of www.ubdt.orgwebsite.

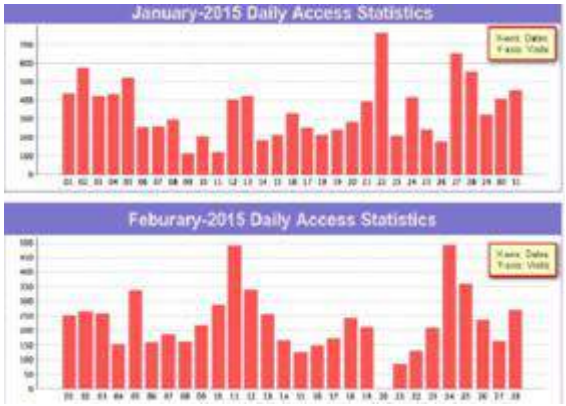


Fig.3 daily access statistics

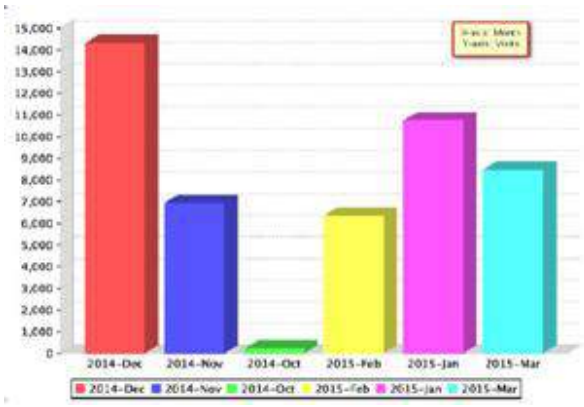


Fig.4 monthly access statistics

Fig.3 indicates more number of visits are on 22nd, 27th, 28th of January and 11th, 12th, 24th February and very less visitors on 9th, 11th of January and 20th of February. Fig.4 indicates more number of visitors are in the month of December and very less visitors in the month of October.

5.3 Access Statistics

This part of our project can be considered as the most important part as it provides which IP produces more hits and more visits and which IP uses high bandwidth. It also helps in determining who all accessed the website. The below table indicates a list of IP addresses that hit the website along with how many times the website was visited by a particular user and how much bandwidth each user used.

Host	Hits	Visitors	Bandwidth(MB)
14.139.152.34	29772	4371	826
216.158.82.218	9391	9262	118
14.139.155.178	1805	143	34
71.198.24.238	1604	93	6
117.241.0.112	1165	214	12
14.141.216.130	1133	180	19
112.133.192.42	1029	150	27
117.240.86.5	811	101	15
37.228.105.7	792	30	7
117.211.56.9	768	208	11

Table 3. Access Statistics

5.4 Visits-per-Country

The table shows Number of visits to the website based on countries.

Country Code	Visits
IN	25465
US	11099
FR	547
CN	297
UA	124
CA	115

Table 4. Visits Per Country

5.5 Errors

The last feature is find what kind of errors people encounter when they access the website. The below chart indicates the errors users encountered while they accessed the website.

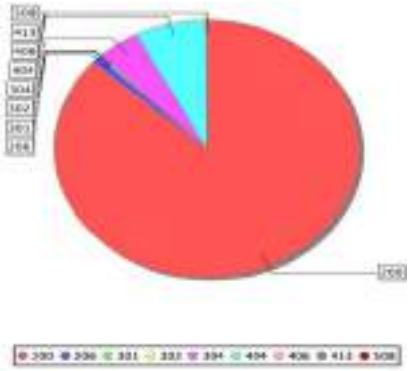


Fig.5 pie chart showing the errors that occur frequently

6.CONCLUSIONS

Web sites are one of the important means for organizations for making advertisements. In order to get outlined results for a specific web site, we need to do log examination that helps enhance the business methodologies and also produce measurable reports. In this project with the help of Hadoop framework web server log files are analyzed. Data gets stored on multiple nodes in a cluster so the access time required is reduced. MapReduce works for large datasets giving efficient results.Using visualization tool for log analysis will give us graphical reports indicating hits for web pages, client's movement, in which part of the web site clients are interested. From these reports business groups can assess what parts of the site need to be enhanced, who are the potential clients, what are the regions from which the site is getting more hits, and so on. This will help organizations plan for future marketing activities. Log analysis can be done using many different techniques however what is important is response time. Hadoop MapReduce model provides parallel distributed processing and reliable data storage for huge volumes of web log files. Hadoop's ability of moving processing to data rather than moving data to processing helps enhance response

REFERENCES.

[1]. SayaleeNarkhede and TriptiBaraskar, "*HMR Log Analyzer: Analyze Web Application Logs over HadoopMapReduce*," International Jour nal of UbiComp (IJU) vol.4, No.3, July 2013.

[2] Liu Zhijing, Wang Bin, (2003) "*Web mining research*", International conference on computationalintelligence and multimedia applications, pp. 84-89.

[3] Yang, Q. and Zhang, H., (2003) "*Web-Log Mining for predictive Caching*", IEEE Trans.Knowledge and Data Eng., 15(4), pp. 1050-1053.

[4] Andrew Pavlo, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden,Michael Stonebraker, (2009)"*A Comparison of Approaches to Large-Scale Data Analysis*", ACMSIGMOD'09.

[5]. Mr.YogeshPingle, VaibhavKohli, ShrutiKamat, NimeshPoladia, (2012)"*Big Data Processingusing Apache Hadoop in Cloud System*", National Conference on Emerging Trends inEngineering & Technology.

[6] Tom White, (2009) "*Hadoop: The Definitive Guide. O'Reilly*", Scbastopol, California.

[7] Jeffrey Dean and Sanjay Ghemawat., (2004) "*MapReduce: Simplified Data Processing on LargeClusters*", Google Research Publication.

[8] Chen-Hau Wang, Ching-Tsorng Tsai, Chia-Chen Fan, Shyan-Ming Yuan, "*A Hadoop Based Weblog Analysis System*", 2014 7th International Conference on Ubi-Media Computing and Workshops.

[9] SayaleeNarkhede and TriptiBaraskar, "*HMR Log Analyzer: Analyze Web Application Logs over HadoopMapReduce*," International Jour nal of UbiComp (IJU) vol.4, No.3, July 2013.

[10] MilindBhandare, VikasNagare et al., "*Generic Log Analyzer Using HadoopMapreduce Framework*," International Journal of Emerging Technology and Advanced Engineering (IJETAE), vol.3, issue 9, September 2013.

[11] Savitha K, Vijaya M S, "*Mining of web server logs in a distributed cluster using big data technologies*", International Journal of Advanced Computer Science and Applications, Vol.5, NO.1, 2014

BIOGRAPHY

Lavanya KS Sis currently pursuing his M. Tech. in Computer Science & Engineering degree from University RRCE College of Engineering, Bangalore, India. She has received B.E degree in Computer Science and Engineering from HMS Institute of Technology, Tumkur under Visvesvaraya Technological University, Belgaum. Her research interests are Big Data Analytics and Web Mining.

Mr.Srnivasa R is working as Associate Professor in Department of studies in CSE, University RRCE College of Engineering, Bangalore, India

Ranking Detection and Avoidance Frauds in Mobile Apps Store

ABHIILASH T P, L DINESHA
PG Scholar, Assistant Professor
Deptartment of Computer Science & Engineering,SSIT,Tumakuru,Karnataka, India
abhilash.panchu@gmail.com , ldinesha.ssit@gmail.com

ABSTRACT- There are millions of apps are available in market for the application of mobile users. However, all the mobile users first prefer high ranked apps when downloading it. But we cannot guarantee the reliability for the downloaded application since there is increasing number of ranking frauds. Ranking fraud in the mobile App market refers to fraudulent or deceptive activities which have a purpose of bumping up the Apps in the popularity list. Indeed, it becomes more and more frequent for App developers to use shady means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. While the importance of preventing ranking fraud has been widely recognized, there is limited understanding and research in this area. To this end, in this paper, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. Specifically, we first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Such leading sessions can be leveraged for detecting the local anomaly instead of global anomaly of App rankings. Furthermore, we investigate three types of evidences, i.e., ranking based evidences, rating based evidences and review based evidences, by modeling Apps' ranking, rating and review behaviors through statistical hypotheses tests. In addition, we propose an optimization based aggregation method to integrate all the evidences for fraud detection. Finally, we evaluate the proposed system with real-world App data collected from the OS App Store for a long time period. In the experiments, we validate the effectiveness of the proposed system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities. There are in huge number of official and unofficial markets are available for mobile users to get variety of application. However, we cannot guarantee that the applications available in the market are trust worthy. Therefore, the application needs to be validated. In this paper we are introducing new protocol for detecting malicious apps.

Keywords— Rank Based evidence, Ranking Fraud, Mobile Applications.

I. INTRODUCTION

Mobile telephone fraud is the unauthorized use of the telecommunications network accomplished via deception. Mobile telephone fraud is a tremendous difficulty for network vendors and their customers: in some local areas it is estimated that more than half the use is fraudulent.

To stimulate the progress of cell Apps, many App outlets launched daily App leader boards, which demonstrate the chart rankings of most trendy Apps. Certainly, the App leader board is likely one of the primary ways for promoting cellular Apps. A bigger rank on the leader board normally leads to a tremendous quantity of downloads and million bucks in earnings. Consequently, App builders are likely to explore various approaches corresponding to promoting campaigns to promote their Apps in an effort to have their Apps ranked as high as possible in such App leader boards.

In the literature, even as there are some associated work, similar to web ranking junk mail detection [1], [2] online overview junk mail detection and mobile App recommendation the main issue of detecting ranking fraud for cellular Apps is still underexplored.

We propose to advance a ranking fraud detection process for mobile Apps. Along this line, we establish a couple of main challenges. First, ranking fraud does now not perpetually happen in the whole life cycle of an App, so we need to discover the time when fraud occurs. Such challenge can be regarded as detecting the local anomaly instead of global anomaly of cell Apps. 2nd, as a result of the big quantity of mobile Apps, it's intricate to manually label ranking fraud for each and every App, so it's fundamental to have a scalable method to mechanically detect ranking fraud without making use of any benchmark information. In the end, due to the dynamic nature of chart rankings, it's not easy to establish and confirm the evidences linked to ranking fraud, which motivates us to observe some implicit fraud patterns of cellular Apps as evidences.

Indeed, our careful commentary displays that cell Apps aren't constantly ranked high within the leader board, however handiest in some leading events, which type distinct leading Sessions.

Ranking fraud most of the time happens in these leading sessions. Therefore, detecting rating fraud of cellular Apps is truly to detect ranking fraud inside leading sessions of cellular Apps. Specifically, we first recommend a easy but powerful algorithm to identify the leading sessions of each and every App established on its historical rating files. Then, with the analysis of Apps' ranking behaviors, we discover that the fraudulent Apps quite often have exclusive ranking patterns in each and every main session when compared with normal Apps. Therefore, we symbolize some fraud evidences from Apps' historical ranking documents, and enhance three services to extract such ranking based fraud evidences. Nonetheless, the rating based evidences can be affected through App builders' fame and some reliable advertising campaigns, reminiscent of "constrained-time discount". As a result, it is not ample to only use ranking established evidences. Consequently, we further suggest two forms of fraud evidences headquartered on Apps' ranking and assessment historical past, which reflect some anomaly patterns from Apps' ancient rating and evaluation documents. Furthermore, we boost an unsupervised

evidence-aggregation system to integrate these three types of evidences for evaluating the credibility of main periods from mobile Apps detection system for mobile Apps.

II. PROPOSED SYSTEM

Figure 1 indicates the block diagram of proposed architecture. For locating the leading session we'd like the ancient data. Old files accrued from quite a lot of sources. Then discovering the leading session shall be finished. This can be completed with the aid of discovering leading events from the App's historic rating documents. Second, we need to merge adjoining main pursuits for setting up leading periods.

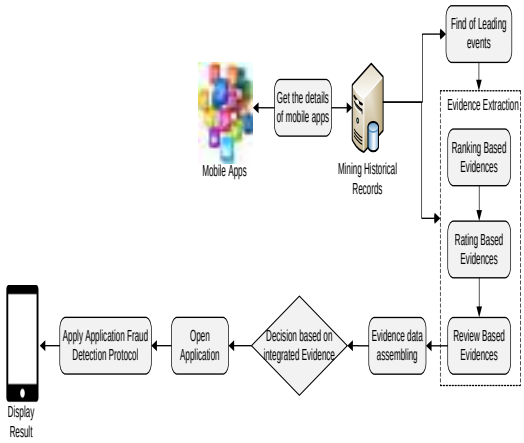


Figure 1: Proposed architectureIn the next step we ought to collect the evidences from the historic documents. Evidences will likely be collected headquartered on prior ranking, rating and experiences respectively.

A. Identifying the Leading Sessions of App's:

a. Preliminaries

The App leader board demonstrates top k popular Apps with admire to specific classes, corresponding to "top Free Apps" and "prime Paid Apps". Additionally, the leader board is mainly up to date periodically. So each and every cellular app 'a' has number of historic files which is denoted with the aid of R (a).

$$R(a) = \{r_1^a, \ldots r_i^a, \ldots . r_n^a\} \quad (1)$$

Where n denotes the quantity of all rating files notice that, the smaller value "ri power a" has the larger ranking function the App obtains.

By means of inspecting the old rating documents of cell Apps, we become aware of that Apps customarily aren't without end ranked excessive within the leader board, however simplest in some leading events. Furthermore, we find that some Apps have several adjoining main hobbies which might be just about each other and kind a leading session. The leading classes of a mobile App characterize its intervals of fame, so the rating manipulation will best take place in these leading sessions. As a result, the trouble of detecting ranking fraud is to detect fraudulent leading sessions.

Consequently, the main issue of detecting rating fraud is to realize fraudulent leading sessions. Along this line, the first task is the way to mine the leading sessions of a mobile App from its historic ranking records.

b. Mining Leading Sessions

There are two fundamental steps for mining leading sessions. First, we have got to observe leading events from the App's historic ranking files. 2nd, we ought to merge adjacent primary event for setting up leading sessions.

c. Extraction of Evidences

In determining ranking frauds historical evidences plays vital role. Here also we collect the useful evidences based on rating, ranking and experiences.

C1.1. Ranking Based Evidences

By way of inspecting the Apps' ancient ranking files, we realize that Apps' rating behaviors in a leading event continuously fulfill a targeted rating pattern, which contains three specific ranking phases, namely, rising phase, maintaining phase and recession phase. Especially, in every leading event, an App's ranking first increases to a top position in the leader board (i.e., rising phase), then keeps such top position for a period (i.e., maintaining phase), and ultimately decreases till the tip of the event (i.e., recession phase).

As soon as a common App is ranked high within the leader board, it obviously owns lots of honest enthusiasts and would appeal to more and more purchasers to down load. Consequently, this App is usually ranked high inside the leader board for a long time. From the above discussion, we propose some ranking established evidences of major sessions to assemble fraud evidences for ranking fraud detection.

Proof 1:

We use two shape parameters θ1 and θ2 to quantify the score patterns of the rising phase and the recession phase of App a's main event e, which may also be computed by using

$$\theta 1^e = arctan \frac{n(K - r_b^a)x^t}{t_b^e - t_a^e} \quad (2)$$

$$\theta 2^e = arctan \frac{(K - r_c^a)}{t_d^e - t_c^e} \quad (3)$$

Where, k is the rating threshold Intuitively, a higher θ1 may just point out that the App has been bumped to a high rank within a short interval, and a tremendous θ2 could factor out that the App has dropped from a excessive rank to the bottom within a brief interval. Hence, a leading session, which has more leading movements with significant θ1 and θ2 values, has larger chance of getting rating fraud. Correct now, we outline a fraud signature θs for a predominant session as follows

$$\theta_s = \frac{1}{|E_s|} \sum_{e \in s} \theta_1^e + \theta_2^e \quad (4)$$

Where $|E_s|$ quantity of primary routine in session s. Intuitively, if a leading session s includes significantly higher θs in evaluation with exceptional leading durations of Apps within the leader board, it has high hazard of having rating fraud. To grab this, we propose to apply statistical speculation scan for computing the value of θs for each major session. In designated, we define two statistical hypotheses as follows and compute the p-price of every leading session.

• The signature θs of leading session s is no longer useful for detecting ranking fraud.

• The signature θs of main session s is enormously better than expectation.

Proof 2:

The Apps with rating fraud most as a rule have a short retaining part with excessive rating positions in each leading occasion. Hence, if we denote the retaining segment of a important party e as $\Delta t_m^e = t_c^e - t_b^e + 1$ and the typical rank on this retaining phase as r_m^e,, we can outline a fraud signature x_s for every main session as follows

$$x_s = \frac{1}{|E_s|}\sum \frac{k^a - r_m^e}{\Delta t_m^e} \quad (5)$$

If a leading session includes greatly higher xs when put next with other leading sessions of Apps inside the leader board, it has high threat of having score fraud. To detect such signatures, we outline two statistical hypotheses as follows to compute the importance of xs for every leading session.

- The signature χs of main session s is not useful for detecting ranking fraud.
- The signature χs of main session s is significantly larger than expectation.

C1.2 Experience based evidences:

Experiences are very predominant proof in deciding whether the data is riskless or no longer. However most likely it's difficult to gauge established on simplest experiences.

Experiences can replicate the individual perceptions and utilization experiences of existing users for targeted cellular Apps. Most likely, comparison manipulation is no doubt probably the most primary perspective of App rating fraud. Mainly, earlier than downloading or purchasing a brand new cell App, users more often than not first of all learns its prior experiences to ease their resolution making, and a cell App entails more positive reports may just attract further customers to download. For that reason, imposters often put up false reviews inside the leading classes of a targeted App so that you could inflate the App downloads, and thus propel the App's ranking positions within the leader board.

Proper here we endorse two fraud evidences based on App's stories behaviors in main sessions for detecting rating fraud.

Proof 1:

Many of the experience manipulations are applied by way of boot farms when you consider that of the immoderate expense of human useful resource. As a consequence, review spammers most of the time submits a couple of duplicate or near-reproduction studies on the equal App to inflate download. In difference, the traditional App perpetually has various studies given that customers have exceptional individual perceptions and utilization. From the above observations, right here we define a fraud signature $\mathrm{Sim}(s)$, which denotes the natural mutual similarity between the reports inside main session s. In particular; this fraud signature can be computed with the help of following steps.

- For each review c in leading session s, we remove all stop words (e.g., "of", "the") and normalize verbs and adjectives (e.g., "plays → play", "better → good").
- We build a normalized words vector $\overrightarrow{wc} = dim(n)$ for each review c, where n indicates the number of normalized words in all reviews of s.

$$\dim[i] = \frac{\mathrm{freq}_{i,c}}{\sum_i \mathrm{freq}_{i,c}} \quad (1 \le i \le n) \quad (6)$$

Freq is the frequency of ith word in c

Finally, we can calculate the similarity between two reviews ci and cj by the Cosine similarity$\cos(\mathrm{Wci}, \mathrm{Wcj})$. Thus the fraud signature $\mathrm{Sim}(s)$ is

$$Sim(s) = 2 * \sum_{1 \le i < j < Nb} cos(Wci, Wcj) \backslash Ns * (Ns - 1) \quad (7)$$

Where Ns is the number of stories for the period of principal session s. Intuitively, the better valued at of sim(s) suggests further reproduction/close-reproduction reviews in s. For that reason, if a leading session has significantly bigger worth of Sim(s) in comparison with one of kind fundamental courses of Apps in the chief board, it has excessive likelihood of having ranking fraud.

To compute this, we define statistical hypotheses to compute the value of Sim(s) for each main session as follows.

- The signature Sim(s) of leading session s is not useful for detecting ranking fraud.
- The signature Sim(s) of leading session s is significantly higher than expectation.

Here, we use the Gaussian approximation to compute the p-worth with the above hypotheses. Above all, we count on $Sim(s)$ follows the Gaussian distribution,

$$\varphi_6^{(s)} = p(\mu_{sim}, \sigma_{sim}) \ge sim(s) \quad (8)$$

Proof 2:

From the real-world observations, we find that each and every overview c is consistently associated with a particular latent discipline z For example; some reviews may just be involving the latent matter "valued at to play" while some may be involving the latent matter "very boring". In the meantime, on account that exceptional purchasers have one-of-a-type private preferences of mobile Apps, every App a would have great subject distributions in their ancient evaluation records. Intuitively, the subject distribution of

studies in a average leading session s of App a, i.e., p/z must be steady with the field distribution in all historic overview files a). It's considering that that the evaluation themes are centered on the customers' private utilization experiences however now not the reputation of cell Apps. In big difference, if the studies of s have been manipulated, the 2 subject distributions possibly markedly particular. For example, there would incorporate extra confident subject matters, similar to "worth to play" and "basic", inside the leading session.

We advise to leverage discipline modeling to extract the latent issues of experiences. Specially, right here we adopt the mainly used Latent Dirichlet Allocation (LDA) mannequin for learning latent semantic issues. To be more special, the historical stories of a cell App a, i.e. Ca, is believed to be generated as follows.

1st, earlier than generating Ca, ok prior conditional distributions of phrases given latent issues ϕz are generated from a previous Dirichlet distribution β.

2nd, a previous latent topic distribution θa is generated from a previous Dirichlet distribution α for each and every cellular App a.

The training procedure of LDA mannequin is to learn right latent variables $\theta = \{P(z|Ca)\}$ and $\phi = \{P(w|z)\}$ for maximizing the posterior distribution of review observations, i.e., $P(Ca|\alpha, \beta, \theta, \phi)$. this paper, we use a Markov chain Monte Carlo process named Gibbs sampling for coaching LDA mannequin. If we denote the experiences in main session s of a as Csa, we are able to use the KL-divergence to estimate the change of topic distributions between Ca and Csa.

$$DKL\ (s||a) = \Sigma\, k\ P(zk|Csa)\ ln\ (\frac{P(zk|Csa)}{P(z_k|C_a)} P(zk|Ca)) \quad (9)$$

Where

$P(zk|Ca))\ and\ P(zk|Cs;a) \propto P(zk)\ \Pi\ w2Csa$ $P(w|zk)$ may also be got by way of the LDA training approach. The larger value of DKL (s||a) suggests the better difference of subject distributions between Ca and Cs; a. For that reason, if a main session has vastly bigger worth of DKL (when put next with other leading sessions of Apps in the chief board, it has high probability of having rating fraud. To seize this, we define statistical hypotheses to compute the significance of DKL (each leading session as follows.

- The signature DKL $(s||a)$ of leading session s is not useful for detecting ranking fraud.
- The signature DKL $(s||a)$ of leading session s is significantly higher than expectation

The Gaussian approximation to compute the p-value with the above hypotheses;

$$\varphi_7(s) = 1 - p\big(N(\mu DL, \sigma DL)\big) \geq DKL(s||a) \quad (10)$$

The values of two evidences Ψ6(s) and Ψ7(s) are in the range of [0, 1]. Meanwhile, the higher evidence value a leading session has, the more chance this session has ranking fraud activities.

d. Proof Aggregation:

After extracting three forms of fraud evidences, the following assignment is the way to mix them for ranking fraud detection. Certainly, there are various rating and proof aggregation approaches in the literature, comparable to permutation established units [7], rating centered items [1], [6] and Dumpster-Shafer principles [1], [2]. However, some of those approaches focus on finding out a worldwide ranking for all candidates. This isn't right for detecting rating fraud for brand new Apps. Different ways are situated on supervised learning strategies, which depend upon the labeled training data and are tough to be exploited. Instead, we advocate an unmonitored method centered on fraud similarity to combine these evidences.

Certainly, we outline the final proof ranking $\Psi(s)$ as a linear combination of all of the existing evidences as observe that, right here we advise to use the linear blend since it has been validated to be robust and is largely used in imperative domains, such as rating aggregation [6], [8].

$$\varphi(s) = \sum_{N_\varphi}^{i} w_i * \varphi_i(s)\ ,s\ .t\ \sum_{i=1}^{N_\varphi} w_i = 1 \quad (11)$$

d. Algorithm for mining leading sessions

Algorithm

Inputs: 1. a's historical ranking record R_{ai}
2. Ranking threshold k
3. Merging threshold Ø

Outputs:
a's leading session Sa

Initialization: Sa = Ø;
$E_{S=\emptyset}$; e =Ø ; S = Ø
$t^e_{start} = 0$;
for each i ∈ [1, $|R_a|$] do
If $r_i^a \leq k$ and $t^e_{start} == 0$ then
$t^e_{start} = t_i$

else if $r_i^a > k$ *and* $t^e_{start} != 0$ then
$t^e_{end} = t_{i-1}$; e =< t^e_{start}, t^e_{end} >;
If $E_S == \emptyset$ then
E_s u = e; $t^s_{end} = t^e_{end}$;
else then
S =< t^s_{start} , t^s_{end}, E_s >;
Sa U= s; S = Ø is a new session
$E_s = \{e\}$; $t^s_{start} = t^e_{start}$;
$t^s_{start} = t^s_{end}$;
$t^s_{start} = 0$; e = Ø
is a new main event
return Sa

III. RESULTS

Figure 2: Home Page

Figure 3: Service Provider Uploading Application

Figure 4: Application List

Figure 5: User Rating and number of downloads

Figure 6: User Comments and Rating when downloading

Application	Pinterest
Fraud Percentage Based on Rating	20%
Fraud Percentage Based on Review	15%
Fraud Percentage Based on Ranking	22%
Result	Good Application

Table 1: Fraud Application Detection Result

IV. Conclusion

Specifically, we first showed that ranking fraud happened in leading periods and provided a process for mining leading periods for each App from its old rating records. Then, identification of ranking based evidences, rating based evidences and experience based evidences for detecting ranking fraud might be accomplished. Additionally we are integrating application fraud detection method to make procedure robust.

REFERENCES

[1]. K. Shi and K. Ali, "Getjar Mobile Application Recommendations with Very Sparse Datasets", International Conference on Knowledge Discovery and Data Mining, 2012.

[2]. N. Spirin and J. Han, "Survey On Web Spam Detection: Principles and Algorithms", SIGKDD Explor, 2012.

[3]. M. N. Volkovs and R. S. Zemel, "A Flexible Generative Model for Preference Aggregation", International Conference on World Wide Web, 2012.

[4]. Clifton Phua, Vincent Lee, Kate Smith and Ross Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research".

[5]. Z.Wu, J.Wu, J. Cao, and D. Tao Hysad, "A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation", International Conference on Knowledge Discovery and Data Mining, 2012.

[6]. S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review Spam Detection via Temporal Pattern Discovery", international conference on Knowledge discovery and data mining, 2012.

[7]. B. Yan and G. Chen, "Appjoy: Personalized Mobile Application Discovery", International Conference on Mobile Systems, Applications, and Services, MobiSys, 2011.

[8]. L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and in precision recall measures", In Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR'03), pages 369–370, 2003.

Stream Processing of Scientific Big Data on Heterogeneous Platforms with Image Analytics on Big Data

Ms. Shruti B Karki, Mr. Auntin Jose. T
PG Scholar, Associate professor
Department of Computer Science, RRCE, VTU, Bengaluru, INDIA
, Department of Computer Science, RRCE, VTU, Bengaluru, INDIA
shruthikarki@gmail.com, auntin123@yahoo.com

ABSTRACT-Image analytics with high performance is a challenge for the processing of big data, as the image data and the video data is a huge amount of big data. This paper will be presented with a case study for image analytics called the parallel connected component labeling(CCL). Generally this is the first step in image analytics. The high performance CCL implementation can be obtained on the heterogeneous platforms, if the suitable parts of the algorithm are processing on a fine grain parallel field programmable gate array (FPGA) along with the multi core processor. The implementation is suitable for the processing of big image and video data in motion which results in the reduction of the amount of memory that is required by the hardware architecture for different image sizes.

Keywords-- Component labeling, Data in motion, Feature extraction, Heterogeneous platform, Image analytic.

I. INTRODUCTION

One of the big data challenges are image analytic tools applicable to photos and surveillance videos. Statistics indicate that 2.5 quintillion bytes are generated every day. Big data is defined based on its main characteristics which are growing in three dimensions: volume, velocity and variety. In this concept, the unstructured data's volume is in the scale of petabytes, and creation of them will in the fraction of the second. Big data in motion is defined as continues data streams at high data transfer rate in the literature. Such kind of big data represent the data sets that cannot be analyzed with conventional algorithms and standard hardware platforms. The processing of the continuously increasing amount of data is done online and locally on the streamed data due to the typical memory capacity and bandwidth limitations, which determine the overall throughput. The new scale of volume, velocity and variety requires the redesigning a number of infrastructure components along with algorithms to support the large-volume, complex and growing data.

Combination of field-programmable gate arrays (FPGAs) and general-purpose computing on graphic processing units (GPGPUs) enable the processing of large scale problems in the field of genomics, graph analytics, social network analytics, bioinformatics etc. This was not possible before, see [5], [6]. For instance, high-performance hybrid core system proposed by Convey Computer which pairs Intel processors with a coprocessor of FPGAs is able to execute data-intensive problems much more effectively [7]. IBM also established Netezza- an advance high performance data analytical tool that has led to an exponential growth in the field of big data analytics. Netezza is based on the IBM blade architecture that uses FPGAs for filtering the input data before data is being processed.

Scientific data are of 4 kinds: raw data, structured data, published data and data linked to the publication [5]. The raw data is generated from observation and experiment of different phenomena. Biological, life sciences and climate generate huge amount of scientific data [5,6]. Usually Images and videos will have the highest amount of volume in scientific data which are analytically prepared to acquire additional value [8], [9]. The preparation is done by extracting different properties of the image such as objects and movements.

Segmentation is the first step in analytic image processing for many video-based applications which is followed by connected-component labeling [10]. Connected component labeling carries the task of labeling all connected image pixels in a binarized image to identify objects or to extract required features of a particular object. The throughput of the CCL can strongly influence the performance of the whole image processing system as it is one of the first complex processing steps in image processing applications. Because of this purpose, a parallel CCL algorithm having memory-efficient architecture is proposed which is suited for high performance image processing applications such as image analytics. The single pass CCL algorithm is memory efficient and therefore mainly suited for FPGAs. It is possible to achieve a high processing throughput by using this proposed architecture and algorithm for performing connected component labeling of streamed images without the necessity of buffering a full image, which will cause a performance limitation either due to limited FPGA-internal memory or due to limited FPGA-external memory bandwidth.

II. RELATED WORK

Connected component labeling (CCL) is the first step in image analytics algorithms. Recently more sophisticated single pass algorithms have gained interest as an alternative to the classical two-pass CCL algorithm for reconfigurable computing. After two scans of the same image, the CCL algorithm will be completed and these have the sequential dependencies. This results in the requirement of high amount of storage and memory for storing full images. To store the full image and labels, a memory with the same size

as the original image is required. The labels are characterized by the place of the pixel within a particular image. If the pixel belongs to the background area of the image a special background label is used, otherwise its label is determined by the labels of adjacent pixels.

Baily et al. proposed FPGA based single pass architectures [12]. There are two drawbacks with the proposed architecture: Its worst case memory requirement is related to the height and width of the image being processed, and also due to the lack of parallelism in which one pixel is processed per clock cycle causing a performance bottleneck in stream processing. Ma et al. reduced the memory requirements [13] of the algorithm significantly described in [14] by reusing the labels. As of this result, the required worst case memory is proportional to the width of the processed image. Kumar et al. proposed a parallel architecture which enhances the single pass algorithm that is used in [15]. The main aim is to store the whole images in prior to processing. In order to have the performance speed-up, the image is divided into different slices, and different CCL units have to process them independently. After this, each of the line from the image slice is fetched sequentially at a time in a round robin manner, and each CCL unit sends a vector to a global FIFO memory which acts as a vector collector. The vector mainly indicates the regions of the processed image slice except the edges. In the next step, a coalescing unit (CU) determines whether two regions of adjacent slices are connected and belong to the same object by processing one or both border of the image slice.

It is necessary to connect the CCL units to the CU to perform the merging operation of neighboring image slices in the round-robin manner. To have a successful merging, each edge region must have a unique, distinguishable label in its image slice. To overcome the limitations of [13] a modified version was proposed [16] which reduces drastically the amount of memory needed for processing by using twokinds of labels, slice local label and slice global labels. In this paper, a CU was introduced that improve the real-time processing of CCL architecture by merging the results of image slices in a memory-efficient way.

III. VIDEO AND IMAGE ANALYTICS FRAMEWORK

Image analytics needs the processing of videos and images to transform pixel level information to object based information for the analysis of certain properties specific for the considered application domain. The domain includes science, industrial measurement techniques or life science applications. Videos are taken as sequence of images with a specific frame rate. The frame rate may reach up to several hundred or thousand frames per second or more for high-speed scientific applications, in such a way that the video data of a single image sensor are in the range of 1 to 10 Gigabytes per second or 0.1 to 1 Petabyte per day. To overcome with this big data in motion, a high performance reconfigurable computing framework is proposed in this section. The framework is composed of a high-speed input data stream and a heterogeneous platform based on high performance reconfigurable computing devices making use of field programmable gate arrays (FPGA) and multi-core CPUs to which the image processing functions are mapped jointly. As shown in Figure 1, the input data stream is connected via several high speed links to the FPGAs that are able to acquire and analyze images as well as videos with a very high frame rate in real-time. Therefore, the framework based on the heterogeneous platform is equipped with integrated image processing capabilities such as segmentation, component labeling and feature extraction.

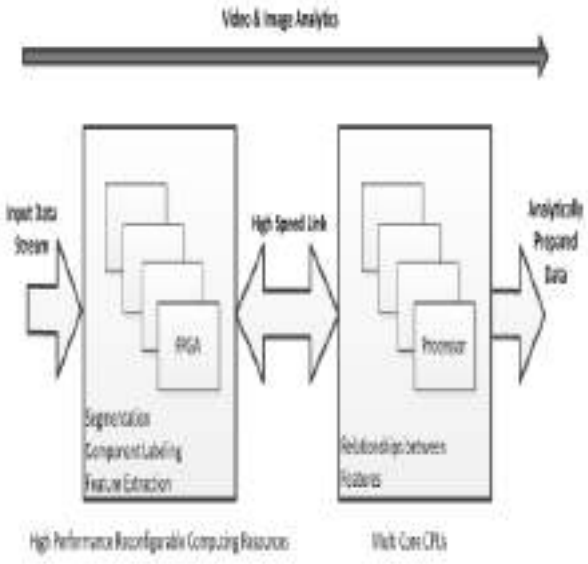


Figure 1. High performance reconfigurable computing framework for processing real-time image stream processing

The image processing steps, which enable a huge data reduction from GBytes of Pixel data to only KBytes of abstract object descriptors - so called feature vectors, are transferred to the multi-core CPU. This reduction allows the framework to output information on every object in every frame even in real-time for very high frame rates. The amount of data which has to be transferred from the FPGA is reduced by several orders of magnitude in this way. Image segmentation and feature extraction are the tasks for image processing architecture realized on the FPGA.

Segmentation is the process of separating the objects from the background when certain threshold is applied. By making use of this method all pixels having an intensity value over a certain threshold are considered as an object and are converted to black. And all pixels below this threshold are converted to white and considered as background. The binary image is generated from the original image accordingly. The major importance in this case is that the threshold value for separating the objects from the background. The method proposed in [17] is used for segmentation by generating a histogram of the captured grayscale image of a scientific application in the field of spray process, as seen in Figure 2. The threshold value is calculated by using the arithmetic mean value of the two

peaks detected in the histogram, which represents the objects and the background.

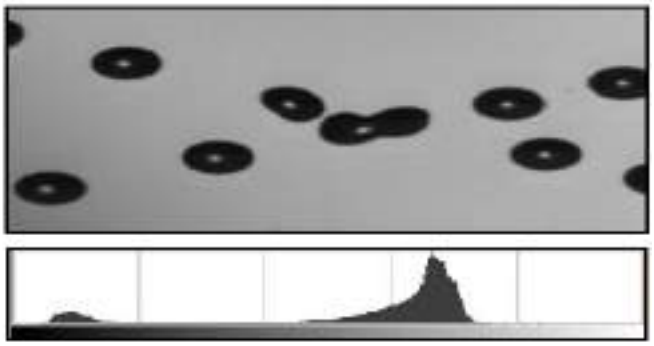


Figure 2. Grayscale raw image taken by image sensor and its histogram.

The next step is feature extraction based on connected component labeling. The main challenge is processing high image frame rate in real-time which requires high bandwidth in the range of 10 to 100 Gbit/s. To overcome this problem, a highly optimized and sophisticated architecture is used. Because of the limited resources in the embedded systems, algorithms which are especially dedicated to FPGA architectures have been proposed [12,13, 14, 16].

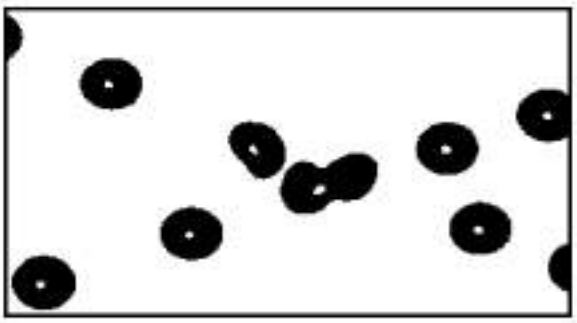


Figure 3. Segmentation of grayscale image with calculated threshold value.

Figure 4 shows the architecture of sequential processing [13] which consists of several components that can be described as follows: the neighborhood context block provides four registers A, B, C and D consisting of four previously processed pixels connected to the present pixel. A, B and C contain the label of the previous row and D contains the previous pixel label. For caching the new labels The row buffer block is used, as they are not saved in a temporary image. The key difference from [11] to [13] is that the labels are reused in every row, and cause the reduction in the need for memory.

As a result of mergers on the previous row, the merger table decides the equality. The translation table modifies a label allocated to the pervious row to the current new label. The label for the current pixel is selected in the label selection block based on the labels of its neighborhood.

The data merging unit records the features of each region by monitoring the labels of the pixels in the neighborhood context block. Each region will have one entry in the data merging unit indexed by the region's label. Whenever a region is updated, even its entry in the data merging unit is also updated.

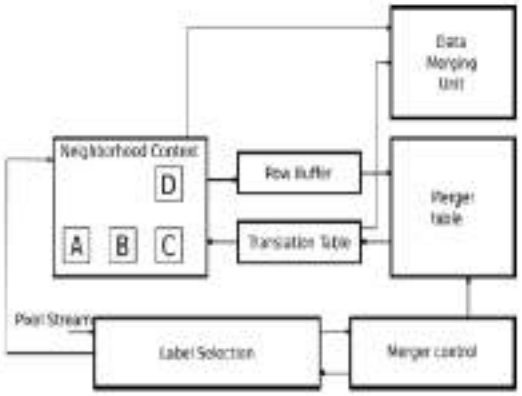


Figure 4. Connected Component LabelingArchitecture proposed in [13]

A bounding box is defined by A and B, two coordinates. Coordinate A indicates the upper left corner and coordinate B indicates the lower right corner. To extract the bounding box for each object the structure and the merging process within the data merging unit is proposed [4 ,6].

To provide bounding box extraction, the data merging unit has to be changed as shown in Figure 5. The modified block has two inputs a and b. One for providing the currently processed pixel's coordinates and the other one for giving information on the neighbor pixels label. The input data is read from the corresponding data table. To find the bounding box for two entries of the data table the following equations are used.

$$x_{1c} = \min(x_{1a}, x_{1b}) \quad (1)$$

$$y_{1c} = \min(y_{1a}, y_{1b}) \quad (2)$$

$$x_{2c} = \max(x_{2a}, y_{2b}) \quad (3)$$

$$y_{2c} = \max(y_{2a}, y_{2b}) \quad (4)$$

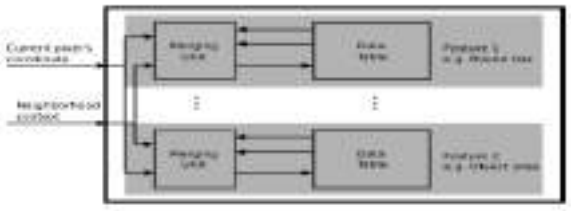


Figure 5. Data Merging Unit for extracting several features of the image objects simultaneously.

A parallelization approach is proposed in order to handle high data throughput. Therefore, the image is divided into different slices, and each slice is processed separately in parallel. In each slices the objects are identified and clubbed by a central unit called the coalescing unit. As a result different pixels are processed simultaneously and speedup based on the number of image slices can be obtained.

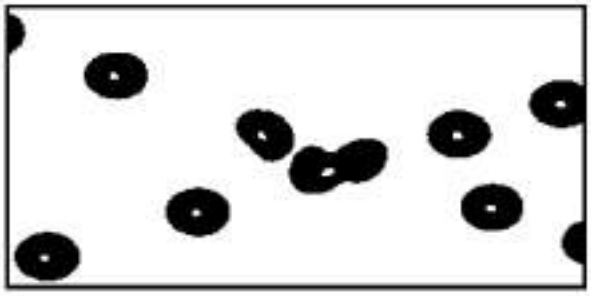


Figure 6. Binarized image after thresholding.

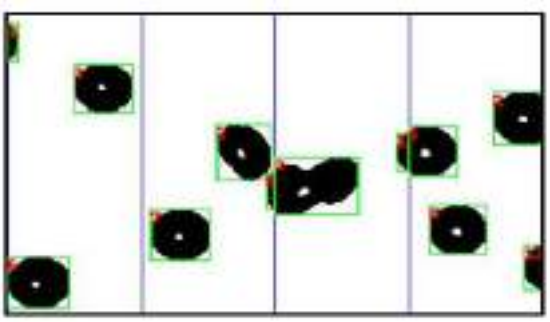


Figure 7. Extracted object features for all sub-images.

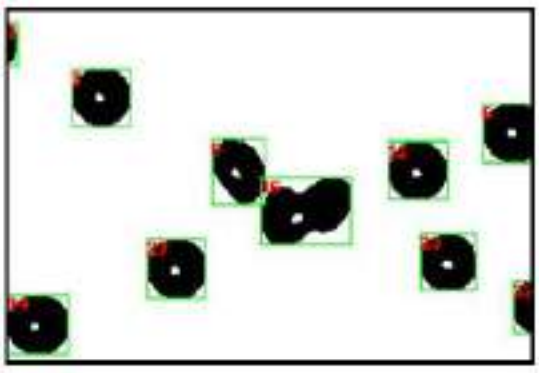


Figure 8. Merged object features for input image.

Figure 6 to Figure 8 show the processing steps for feature extraction. Binary image after the segmentation is shown in Figure 6. In the next step, the image is divided into many image slices for processing in CCL processing units that work in parallel. The bounding box is extracted by each CCL proceeding unit, as it is shown in Figure 7. The bounding boxes touching the slice borders will be merged together to have the correct result which is the last step. This step is depicted in Figure 8.

IV. EXPERIMENTAL RESULTS

The experimental data to examine the potential of heterogeneous systems including reconfigurable logic devices and general purpose processors (GPP) in the context of an image processing system for big data analytics is given. For this examination the approaches from [14], [11] and [16] are evaluated on both reconfigurable logic devices and in software on a general purpose processor. The approach in [16] gains a speedup from dividing the image in different vertical slices using the architecture from [14] as slice processors and merging the results using a coalescing unit. The results for the implementation on a single FPGA where parallelism is inherently used and in software for different image sizes on a single core are given in Table 1.

For the implementation of an array of slice processing units (SPU) having up to 100 SPUS on a single FPGA was realized, in which all are processing an individual image parallely. For the software implementation the classical two-pass algorithm [11] is used to process one image on a single core of a general purpose processor (GPP).

TABLE I. CCL-BENCHMARK CPU VS. FPGA

		Hardware Platform				
Image size		CPU		FPGA		Result
Width	Height	Gigapixels/s	Cores #	Gigapixels/s	Area %	Speedup
128	128	0,063011	1	15,3	8	242
256	256	0,06277	1	11,3	43	182
512	512	0,06286	1	12,3	81	198
1024	512	0,06189	1	9,1	93	146
2048	1024	0,06062	1	6,9	83	114

When comparing the throughput of the software implementation on a GPP and the dedicated architecture on an FPGA, the bandwidth of the FPGA architecture for CCL is one to two orders of magnitude higher. The hardware architecture is highly optimized to the FPGA structure, while for the presented software solution further research has to be done to make a proper comparison, but no decrease in speedup by less than one order of magnitude can be expected. This allows the FPGA architecture to process an image stream consisting of several different image slices simultaneously in real-time. For the case of processing a single image, only one slice processing unit of the SPU array can be used. Still the FPGA architecture accelerates the processing by approximately a factor of 2 compared to a single GPP core. By using the parallelization scheme from [16], several slice processing units can be used to process a single image in parallel enabling a higher throughput. Figure 9 through Figure 11 show the results for this approach where the coalescing unit is realized for FPGAs. Depending on the image size up to 4.5 GPixels/s can be processed. To study the performance of the coalescing process on a GPP, a prototype implementation of a software version of the coalescing unit was realized. The result for the throughput given in Table 2 can be achieved for the case that the feature vectors provided by the slice processing units are already

stored in the systems RAM and can be accessed at full memory bandwidth.

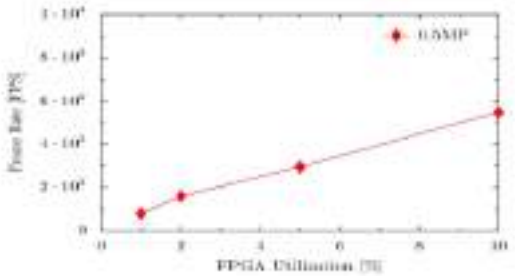


Figure 9. Achievable frame rate for an image with 0.5 Megapixels.

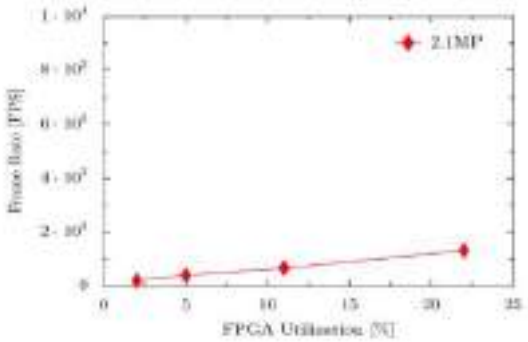


Figure 10. Achievable frame rate for an image with 2.1 Megapixels.

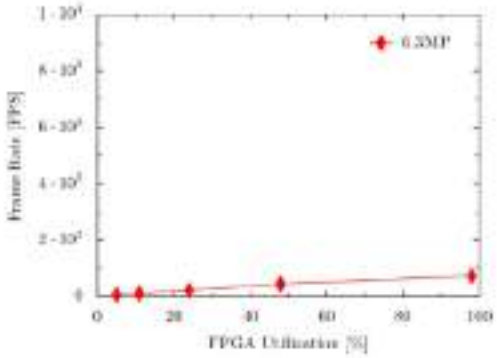


Figure 11. Achievable frame rate of maximum 800 fps for an image with 6.3 MegaPixels and a throughput of 5 Gigapixels per second.

TABLE II. SOFTWARE COALESCING:

Image size / # of slices	*Throughput (Gigapixels/s)*
256 × 256 / 4	6,62
1024 × 1024 / 8	10,51
2048 × 1024 / 4	13,40

THROUGHPUT IN GIGAPIXEL/SEC

Under these conditions the coalescing unit on the GPP achieves up to 13 GPixels/s, which is in the same order of magnitude as the FPGA implementation. In general, the irregular data structures and the sequential algorithm of the coalescing process are better suited for a GPP software implementation. Considering the performance results of both slice processing units and the coalescing unit on a heterogeneous platform, the conclusion is that the slice processing units should be implemented on a FPGAs and the coalescing unit should be implemented on a GPP in order to achieve maximum performance for big data image analytics. Achieving a throughput of 5 Gigapixel/s is equivalent to 0.4 Petabytes per day, so the implementation on a heterogeneous platform with a single mid-sized FPGA can be scaled up linearly with a several FPGAs and processor cores under the condition that several different images are processed in parallel. Beyond these performance considerations, mapping the irregular data structure and the sequential part of the algorithm, which is the coalescing algorithm, to the GPP is of advantage because of its higher processing frequency. For the regular data structures used in the slice processing units carrying out the parallel part of the algorithm, the FPGA achieves a higher throughput due to its parallel architecture.

V. CONCLUSION

Images and videos have the highest amount of volume among big data and therefore high performance image processing plays a very important role for image analytics. In this paper, we have investigated a parallel component labeling algorithm including feature extraction with a broad set of features as an important part of an image analytics framework. It is shown here that the performance of the proposed parallel component labeling algorithm with generalized feature extraction is accelerated and optimized if it is mapped to and executed on a heterogeneous hardware platform based on a fine-grained field-programmable gate array and a multi-core processor. For the parallelized connected component labeling (CCL) algorithm of this paper, the required memory compared to the requirement of algorithms known in the literature is reduced significantly on a heterogeneous hardware platform for typical image sizes even when compared to similar sliced parallel single pass CCL algorithms and architectures such that memory size nor memory bandwidth of the hardware platform has an impact on the performance. The basic structure of the algorithm and architecture is a set of parallel CCL units generating feature data of image slices being coalesced in a separate and subsequent coalescing unit. In order to achieve highest performance, it was shown that the parallel CCL units should be implemented on fine-grained FPGAs and the coalescing unit should be implemented in software on a multi-core processor. With the achieved throughput of 5 Gigapixels per second or 0.4 Petabytes per day the implementation on such a heterogeneous platform with a single mid-sized FPGA can be scaled up linearly with a plurality of FPGAs. Thus, the concept has been proven to be ideally suited for high performance image analytics for big data in motion.

REFERENCES

[1] U. Chandrasekhar, A. Reddy, and R. Rath, "A comparative study of enterprise and open source big data analytical tools," *Conference on Information*

Communication Technologies (ICT), 2013 IEEE, pp.372–377.

[2] T. Eaton, D Deroos, "Understanding big data,"in McGraw-Hill Companies, April 2012.

[3] C. P. Arkady B. Zaslavsky and D. Georgakopoulos, "Sensing as aservice and big data," *in Proceedings of the International Conference on Advances in Cloud Computing (ACC)*, July 2012.

[4] Hirzel, M.; Andrade, H.; Gedik, B.; Jacques-Silva, G.; Khandekar, R.; Kumar, V.; Mendell, M.; Nasgaard, H.; Schneider, S.; Soule, R.; Wu, K.-L., "IBM Streams Processing Language: Analyzing Big Data in motion," *IBM Journal of Research and Development* , vol.57, no.3/4, pp.7:1,7:11, May-July 2013

[5] Y. Demchenko, Z. Zhao, P. Grosso, A. Wibisono, and C. de Laat,"Addressing big data challenges for scientific data infrastructure," *4th International Conference on Cloud Computing Technology and Science (CloudCom)*, 2012 IEEE, pp. 614–617.

[6] S. H. Ketan Paranjape and B. Masson, "Heterogeneous computing inthe cloud: Crunching big data and democratizing hpc access for the life sciences," in Intel White Paper.

[7] C. Computer, "Hybrid-core: the big data computing architecture," in Convey White Paper.

[8] D. Garlasu, V. Sandulescu, I. Halcu, G. Neculoiu, O. Grigoriu, M. Marinescu, and V. Marinescu, "A big data implementation based on grid computing," *in Roedunet International Conference (RoEduNet), 2013 11th*, pp. 1–4.

[9] Oracle, "Big data for the enterprise," in white Paper, 2013.

[10] K. Appiah, A. Hunter, F. Meng, and P. Dickinson, "Accelerated hardware video object segmentation: From foreground detection to connected components labelling," *in Computer vision and image understanding*, 2013, pp. 1282–1291.

[11] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," J.ACM, vol. 13, pp. 471–494, October 1966.

[12] D. Bailey and C. Johnston, "Single pass connected components analysis," *in Proceedings of Image and Vision Computing New Zealand* 2007, dec. 2007, pp. 282–287.

[13] N. Ma, D. Bailey, and C. Johnston, "Optimised single pass connected components analysis," *in ICECE Technology, 2008. FPT 2008. International Conference on*, dec. 2008, pp. 185 –192.

[14] D. Bailey, C. Johnston, and N. Ma, "Connected components analysis of streamed images," *International Conference on Field Programmable Logic and Applications*, sept. 2008, pp. 679 –682.

[15] V. Kumar, K. Irick, A. Al Maashri, and N. Vijaykrishnan, "A scalable bandwidth aware architecture for connected component labeling," *in VLSI ISVLSI), IEEE Computer Society Annual Symposium on*, july 2010, pp. 116 –121.

[16] M. Klaiber; L. Rockstroh; Z. Wang; Y. Baroud , S. Simon "A Memory-Efficient Parallel Single Pass Architecture for Connected Component Labeling of Streamed Images" *International Conference on Field-Programmable Technology (FPT)*, Seoul, Korea, Dec.2012.

[17] R. C. Gonzalez and R. E. Woods. Digital Image Processing (3rd Edition). Prentice Hall, August 2007, pp.741-76

Analysis of DWT OFDM using Rician Channel and Comparison with ANN based OFDM

Geeta S H1, Smitha B2, Shruthi G, Shilpa S G4
Department of Computer Science and Engineering, DBIT, Bangalore,
Visvesvaraya Technology University,Belgaum, Karnataka, India.
1blh_geeta@hotmail.com, 2smithabgowda6@gmail.com,
3shruthigmysore@gmail.com,4shilpashri.06@gmail.com

ABSTRACT- As the generation grown up, designing of the wireless communication system towards providing best performance by the services such as high speed of data delivery, high capacity and high bit rate. MC-CDMA (Multi Carrier-Code Division Multiple Access) is a desired technique for such cases, MC-CDMA gives best results by the combination of OFDM (Orthogonal Frequency Division Multiplexing) and CDMA (Code Division Multiple Access). So far OFDM used FFT (Fast Fourier Transform) as a basic building block, due to the intolerable cyclic prefixes OFDM led to the replacement of FFT by suitable transform. In this paper performance of OFDM can be increased by replacing FFT based OFDM by DWT based OFDM with Rician fading channel, analysis can be done using Linear Analysis which reduces the time complexity, increases the bandwidth efficiency and decreases the complexity. The spreading code like Walsh code is used in receiver and transmitter side for spreading the data. Also doing the comparison of DWT based OFDM with ANN against DWT based OFDM with Rician Fading Channel using performance of BER (Bit Error Rate).

Keywords- ANN, DWT, FFT, OFDM, Rician Fading Channel.

1. INTRODUCTION

MC-CDMA is a combination of OFDM and CDMA, allows the system to support multiple user at the same time using same frequency [3]. Widely used in video broadcasting, wireless LAN's, digital audio, fixed wireless access in broadband and in cellular mobile Communication. OFDM is a technique in wireless digital communication and it is an encoding method of digital data on multiple carrier frequencies. In high data rate applications OFDM has been used successfully, Inter symbol interference problem can be easily solved, eliminates the fading effects and enhance the capacity of channel. The combination of both OFDM and CDMA maintains the different data rates and increase the efficiency of bandwidth.

The primary work is based on FFT in which OFDM systems are intolerable due to the generation of cyclic prefixes, results in loss of bandwidth and consumes more time [2]. In Existing system the FFT based OFDM systems are replaced by DWT based, using ANN channel estimation using Levenberg-Marquardt training algorithm using AWGN channel and Rayleigh channel [1]. In proposed system ANN channel is replaced by Rician channel and do the analysis using linear analysis Results in Reduces the time complexity, increases the bandwidth efficiency and increases the performance. Also Comparison of FFT based OFDM and DWT based OFDM and using ANN based Channel and Rician Channel is included in this Paper. There are three modulation techniques in analog communication are amplitude modulation, frequency modulation and phase modulation. In this paper using frequency modulation to achieve the high data rate transmission to increase the performance. In frequency modulation there are many modulation techniques are there like BPSK, QPSK and QAM. Among these three BPSK modulation technique is used, which gives better result than other techniques.
Also doing the comparison of DWT based OFDAM against FFT based OFDM using performance of Bit Error Rate(BER) [4].

This paper is framed by as follows: Section II represents the spreading using Walsh code, Section III describes FFT based OFDM, Section IV describes the DWT based OFDM, Section V describes DWT based OFDM using ANN channel estimation, Section VI Proposed DWT based OFDM using Rician Channel and analyzed using Linear Analysis, Section VII comparison and simulation results, finally Section VII concludes the main outcome of this paper.

II. SPREADING USING WALSH CODE

[5]Spreading is a transmission in which signal occupies the wider bandwidth than the required minimum bandwidth for information transmission. In MC-CDMA spreading codes are major elements which are combined with the transmitted data stream increases the bandwidth requirement. Spreading provides lots of benefits by encoding the transmitted signal. For each information signal using of spreading codes results the co-existence of multiple coded channels at the same time over same frequency. Spreading signals over a wider bandwidth allow its co-existence with narrow band signals. There are two major spreading factors are direct sequence and frequency hopping [4].

Walsh codes are used in direct sequence spread spectrum (DSSS). It is an decodable code which decode the original message with high probability, within a small fraction of received word. Most important property of Walsh code is Orthogonality, provides Zero cross-correlation between any two Walsh-Hadamard codes when the system is

synchronized. Over a binary alphabet Walsh code is a linear code, maps the message of length n to codewords of length 2^n.

Walsh codes are generated by Hadamard matrix with

$$H1=0 \qquad (1)$$

Where H1 is 1 X 1 matrix with an order 1. The Hadamard matrix is generated by

$$H_{2n} = \begin{bmatrix} H_n & H_n \\ H_n & \overline{H_n} \end{bmatrix} \qquad (2)$$

Hadamard matrix

$$H_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \qquad (3)$$

Hadamard matrix for 4 will be:

$$H_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad (4)$$

Binary data can be usually map to polar form. 0s are mapped to 1s and 1s are mapped to -1.

$$H_1 = 1 \qquad (5)$$

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad (6)$$

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \qquad (7)$$

III. FFT BASED OFDM

FFT based transceiver block diagram is shown in the Fig. 1. The digital input data are processed by using M-ary QAM or PSK modulator to mapping of data with N subcarriers and are implemented using IFFT block[6].

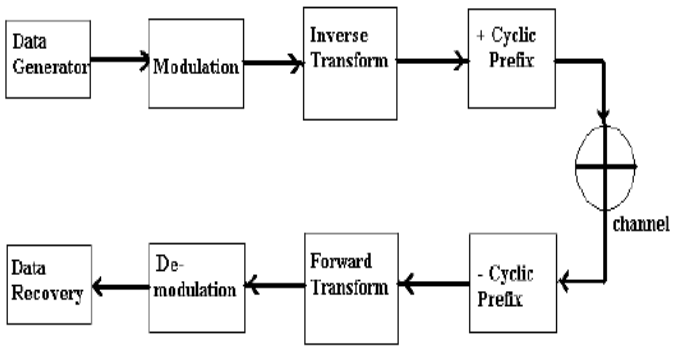


Figure 1. Basic block diagram of FFT based OFDM [2]

Initially the input data is mapped using any of the modulation techniques like BPSK, QPSK and QAM, after the mapping of input data convert the mapped input data into parallel form which is necessary. Each stream of parallel data represents a sub-channel, since serial to parallel converter is used. To modulate the converted data IFFT block is used for stream of low data rate.

The IFFT output is the sum of information signals in the domain of discrete time as following:

$$x_k = \frac{1}{N}\sum_{m=0}^{N-1} X_m e^{j2\pi kn/N} \qquad (8)$$

Where $\{xk|0 \le k \le N-1\}$ is the sequence of discrete time domain, $\{Xm|0 \le m \le N-1\}$ these are complex numbers in the discrete frequency domain using applications of digital modulation method can obtained easily. Applying IFFT to the channels cyclic prefixes will generate, these cyclic prefixes are solved by using Inter Symbol Interference and Inter Carrier Interference [2]. Obtained digital data is then converted to serial form and serial data are transmitted using channel. The receiver receives the data and converted it to serial to parallel form then apply the FFT, using spreading code data are decoded and the final output will be the sum of received signal in discrete frequency domain is as follows:

$$X_m = \sum_{k=0}^{N-1} x_k e^{-\frac{j2\pi km}{N}} \qquad (9)$$

Cyclic prefixes in FFT are used to remove the ISI, results in the cause of overhead, these overheads are sometimes more effective to the systems. The block diagram of FFT based OFDM is shown in the Figure 1. The DWT based OFDM is explained in the Next Section.

IV. DWT BASED OFDM

In orthogonality the Discrete Wavelet Transform (DWT) exhibits best results by reducing the complexity of ISI and ICI problem of cyclic prefixes in FFT based OFDM. The CDMA system requires no need of cyclic prefixes which is overcome by using DWT instead of FFT. The block diagram of DWT based OFDM is shown in Fig. 2.

Wavelet transform tool is a combination of time-frequency, which provides information and representation of time and frequency of a signal simultaneously. The usage of wavelet in OFDM reduces FFT based ISI and ICI problem without the use of cyclic prefixes. In the modulation and demodulation transceiver of FFT based OFDM systems are replaced by Inverse Fast Fourier Transform (IFFT) to Inverse Discrete Wavelet

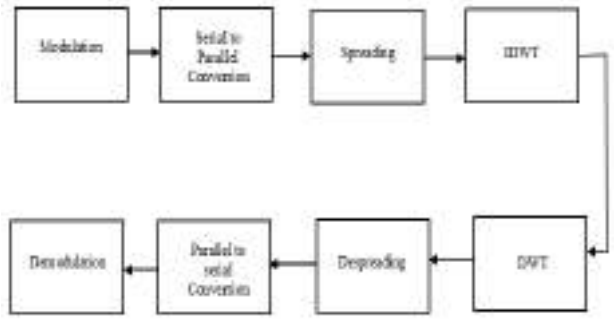


Figure 2. DWT based OFDM

Transform (IDWT) and Fast Fourier Transform (FFT) to Discrete Wavelet Transform (DWT). The outcome of IDWT is represented as follows:

$$d(k) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} D_m^n 2^{m/2} \psi(2_k^m - n) \qquad (10)$$

The above equation is an combined Wavelet co-efficient and Wavelet function with compressed factor m times and shifted n times. For each subcarriers [2]. The process is reversed at the receiver side the output of DWT is as follows:

$$D_m^n = \sum_{k=0}^{N-1} d(k) 2^{m/2} \psi(2_k^m - n) \qquad (11)$$

The random binary data is generated at the receiver end, in which received data are modulated using any of the modulation techniques [7]. The ANN based channel estimation is used, which provide the better use of channel state information. The proposed DWT based OFDM using ANN is described in the next section.

V. DWT OFDM USING ANN BASED CHANNEL ESTIMATION

[1] In wireless systems MC-CDMA are optional for high data rate transmission, which is obtained by the combination of OFDM and CDMA systems are considered as an option in wireless system due to the random probability distribution nature and due to some of the existing challenges. A technique called Artificial Nueral Network (ANN) is a learning based tool used for such a cases, are also a thing that is or may be chosen. ANN are preferred for the better use of channel state information. Here based on the Levenberg-Marquardt (LM) training algorithm ANN is used for channel estimation for Multi-antenna setup over a different channel models in MC-CDMA systems.

Feed Forward (FF) mode is used in ANN is trained under multiple channel states. This approach reduces the complexity of the proposed system and no need of matrix computation. Replacing of FFT with DWT increases the performance. Uses FFT which results in the loss of bandwidth, no efficient transmission and not able to manage the different data rates. So here it deals with the Discrete Wavelet Transform (DWT) based application using ANN based chnel estimation instead of Fast Fourier Transform (FFT) due to the attractive properties of DWT.

5.1 Artificial Neural Network (ANN)

In Artificial Intelligence ANN is one of the popular branch. Consists neuron like processing elements. These neuron like elements are connected each other using weight. Weight is adjusted dynamically till the desired output generation for a desired input.The learning based tool like ANN is interconnected artificial neurons packed in multiple layer among these input is the first one and output is the last one in t of r

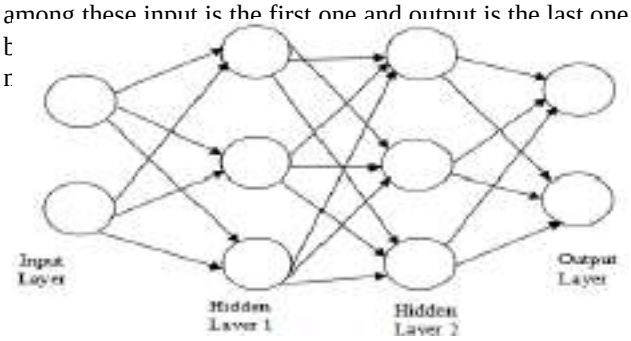


Figure 3. Multi-layer Feed Forward ANN

For the information processing neurons use the linear or nonlinear activation functions. Combined response is obtained by the togetherness of different layers. The difference between the layers and their weights are updated till the expected goal is reached and is called training. ANN in FF mode using Back-Propagation algorithm for training purpose, after training is completed ANN is used for testing purpose [8][9]. The input- output relation of multi-layer FF mode is given as follows:

$$y_k = \phi(\sum_{j=0}^{m} w_{k,j} x_j + b_k). \qquad (12)$$

Where X_j is the input vector, $W_{k,j}$ is the weight vector, b_k is the vector of biases and Y_k is the output vector. The operator *(o)* is the element-by-element activation function for each neuron [3].

5.2. System Architecture for DWT based OFDM using ANN Channel Estimation

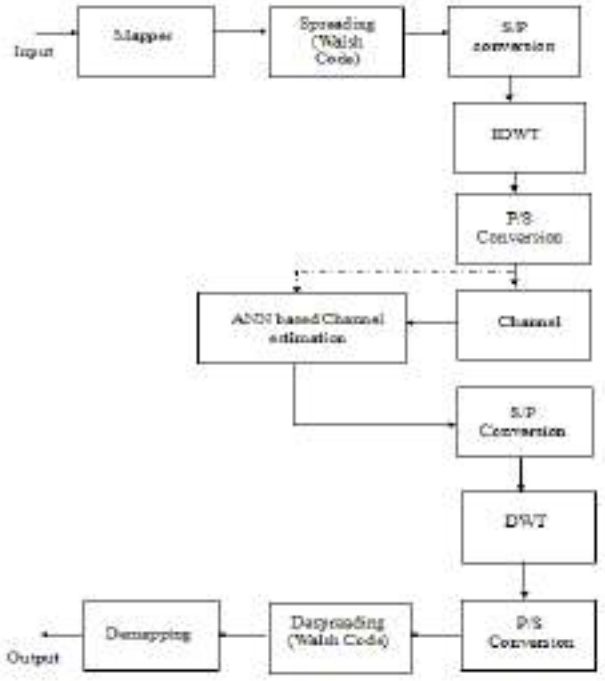


Figure 4. DWT based OFDM using ANN Channel

[8] In the above architecture ANN channel is used for the channel estimation in DWT based OFDM results in the performance increase. Here the system starts from the Mapper which take the bits as input and map the bits using BPSK modulation technique, spreading the modulated data using Walsh Code and converted encrypted data into Serial to Parallel form, apply the IDWT for the obtained data, again convert that data into Parallel to serial, pass those data through the ANN channel as well as Channel and estimate the performance in the ANN channel and reverse the above procedure for the output. Shown in Fig. 4. This estimation can be implement in the matlab ANN tool. In existing system based on the results and analysis it requires more time, which means that increases the time complexity. It is overcome in the proposed system using Rician Channel, results in the reduces the time complexity and increases the performance. Analyzed the result using Linear analysis.

VI. PROPOSED DWT BASED OFDM USING RICIAN CHANNEL AND ANALYZED USING LINEAR ANALYSIS

The related work is for the design of DWT based MC-CDMA in OFDM using Rician Channel for the best performance and to reduces the time complexity faced in the existing system. The below Fig. 5 shows the architecture of DWT based OFDM using Rician Channel and using Linear analysis to analyze the obtained result.

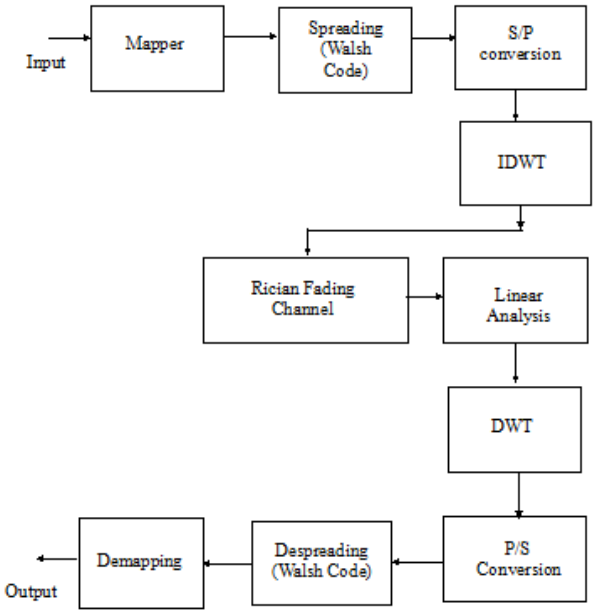


Figure 5. DWT based OFDM using Rician Channel with Linear Analysis

In proposed system MC-CDMA gives the best performance with efficient time by making use of the Rician fading channel and linear analysis. Coming to the working of above architecture, starts with the taking input, it may be any kind of data in this implementation taking bits as an input value and mapping the input using mapper. Mapper includes many techniques used for mapping the data, those techniques are BPSK, QPSK, QAM and etc..among those techniques BPSK used in our implementation for the best results and the mapped data is spreaded using spreading techniques, there are three spreading techniques like Walsh code, PN sequence and gold codes. In the above architecture Walsh code is used for spreading which is nothing but encoding technique and converting the spreading data into serial to parallel transform, for the parallel form of data apply the IDWT transformation technique, the IDWT wavelets are passed to the Rician fading channel which is for the multipath communication with line-of-sight, this Rician channel provides best frequency range led to increase in the performance of channel capacity, transmission range and reduces the time complexity. After estimation in Rician channel give the result to the linear analysis, analyze the result and gives the best linear result for the obtained data. These process happened in the transmission part and same process is reversed in the receiver part for the output data. In the receiver part for analyzed data apply the DWT technique and converting the wavelet into parallel to serial form and despreading is applied, finally demapping using demapper. The result is best performance while transmission, channel capacity reduces the complexity of time and obtain the fading result using BER. This simulation can be done in matlab using Linear analysis tool.

VII. COMPARISON AND SIMULATION RESULT

The comparison and simulation results are executed in proposed system using MATLAB. Fig. 6 shows that mutual information of Rayleigh channel obtained using ANN channel estimation over the SNR to bit error rate, compared using conventional pilot based estimation, ANN estimation (SISO) and ANN estimation (MIMO). Fig. 7 shows the Frequency Response in proposed system using Rician Channel. Fig. 8 gives frequency response in running state using Rician fading channel compared against frequency and magnitude. Fig. 9 shows that Eb/No Vs BER for BPSK over Rician Fading channel with AWGN noise, results in best as compared to the existing system. Fig. 10 shows the Fading Result over BER and SNR using Rician Fading channel. Fig. 11 shows the frequency response estimation for the result obtained in Rician channel over frequency, phase and magnitude. Fig. 12 shows the Linearization results for the obtained frequency estimated result for the linear display of the obtained result against time and amplitude. Based on the simulation result Rician channel is a best channel for transmission in the multipath communication majorly used in mobile communication.

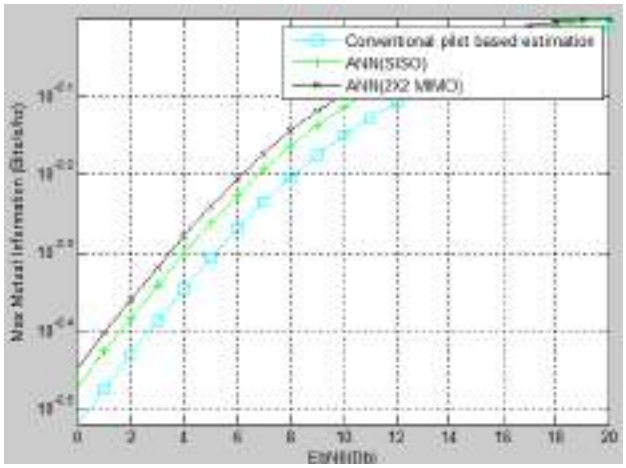


Figure 6. Mutual information in Rayleigh channel using ANN

Figure 7. Frequency Response in proposed system using Rician Channel

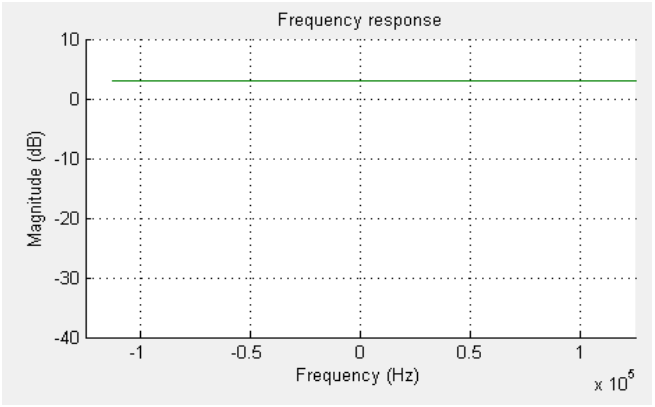


Figure 8. Frequency response in running state

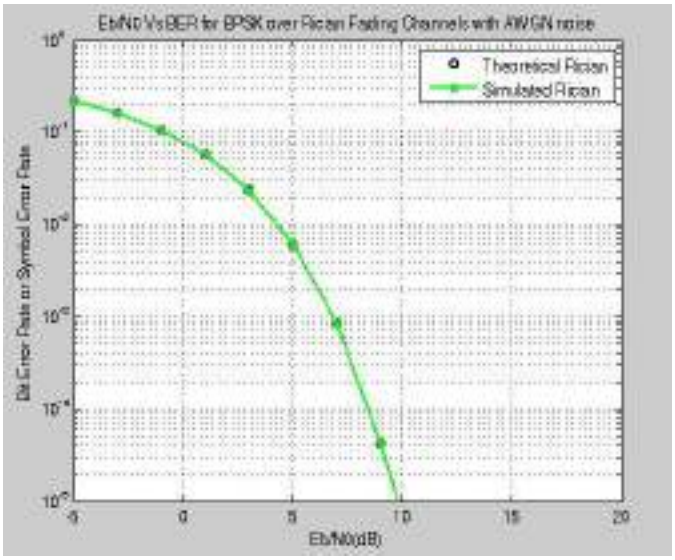


Figure 9. SNR Vs BER over Rician Channel

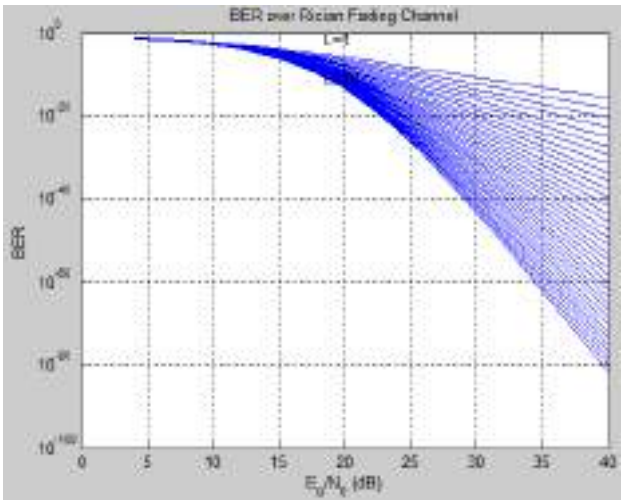


Figure 10. BER over Rician Fading Channel

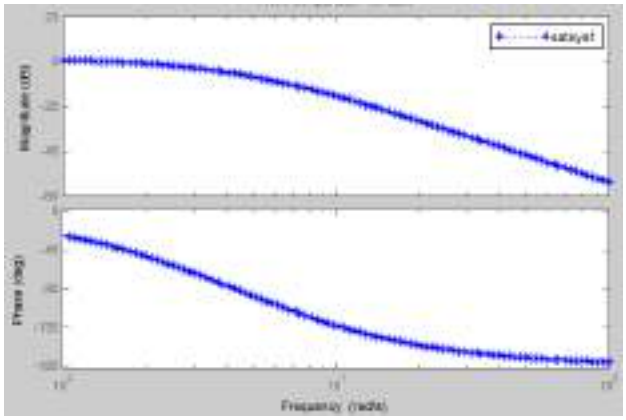


Figure 11. Frequency Response Estimation

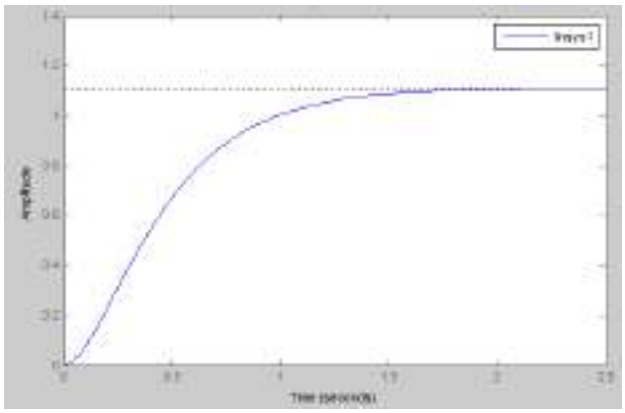


Figure 12. Linearization for Frequency Response Estimation

CONCLUSION

As per the above work studies the DWT based OFDM using ANN and DWT based OFDM using Rician Channel with Linear analysis. The disadvantage of FFT is cyclic prefixes is overcome by using the DWT based OFDM, pilot carrier is removed in the proposed system, BPSK modulation technique is used for the mapping, Walsh code is used for spreading the mapped data, applied IDWT for spreading result and passes through the Rician Channel which is an multipath channel communication results in the best performance, decreases the time complexity, increases the speed of transmission, utilizes the lesser bandwidth and linear analysis tool is used for the linear analysis for the obtained result. Finally conclude these result encourages the further implementation using DWT in the wireless application.

REFERENCES

[1] Mayurakshi Roy Medhi and Kandarpa Kumar Sarma, "DWT based Multi Carrier CDMA System using ANN-Aided Channel Estimation" 2nd International Conference on Signal Processing and Integrated Networks(SPIN).

[2] Miss Krupali N. Umaria and Prof. Ketki Joshi, "Comparative Analysis of BER Performance of DWT based OFDM System with Conventional FFT Based OFDM System".

[3] Shinsuke Hara and Osaka University Ranjee Prasad, Delft University of technology "Overview of MC-CDMA" IEEE Communication Magazine.

[4] I.Sowmya and J.Jayakumari, NICHE "Simulation and Analysis of MC-CDMA System Using Discrete Wavelet Transform".

[5] Saleh Masum, M.Hasnat, Md.Matiqual Islam, Rifat Ava Shams and Shaikh analyst "Impact of Different Spreading codes using FEC on DWT Based MC-CDMA system".

[6] Swati Sharma and Sanjeev Kumar, "BER Performance Evaluation of FFT-OFDM and DWT-OFDM", International journal of Network and Mobile technologies.

[7] D.K Sharma, A Mishra & Rajiv Saxena, "Analog and Digital Modulation techniques: An Overview " TECHNIA International Journal of Computiong Science and Communication technologies.

[8] Cebrial Ciflkli A.Tuncay Ozsahin.A.Cagri Yapici " Artificial neural network channel Estimation based on Levenberg Marquardt for OFDM Systems".

[9] Kanchan Sharma, Shwetha Varshney Dept of Electronics and Communication technology, guru gobind Singh Indraprastha University, Delhi, India" Artificial Neural Network Channel Estimation fot OFDM System".

Automatic range timeline generation for volume based Tweet stream

1Tejaswini K,2Madhavi H,3Sreenivasa B R
Dept. of CSE, Rajarajeshwari College of Engineering, Bangaluru, India
1tejaswini.krish@gmail.com, 2manasamadu@gmail.com,3br.sreenu@gmail.com

ABSTRACT-Now a days twitter is used worldwide millions of users share and create the tweets. While this is informative, can also be beyond limits even though tweets is in raw form. Since millions tweets are shared and created it is causing lots of collision. It is very difficult job to handle. Sumblr is a proposed for novel continuous summarization framework called to eliminate the problem. Which is designed to deal with dynamic, quick arriving, and large form tweet streams, on the other hand traditional document summarization methods which focus on static and small form data set. Our proposed framework consists of three major phases. To cluster tweets and maintain distilled enumeration in a data structure known as TCV (tweet cluster vector), we propose an online tweet stream clustering algorithm. We proposed a TCV-Rank summarization technique for creating online summaries and historical summaries of arbitrary time durations.

Keywords—Tweet stream, sumblr, online/historical summary, timeline, evolutionary, TCV.

I. INTRODUCTION

With the Rise of popularity in microblogging services such as Twitter, Myspace and Facebook in the explosion of the amount of short-text messages. Twitter, for instance, which receives over approx. 400 million tweets over a day has emerged as an overwhelming source of news, blogs, opinions, and lot more. Twitter may yield millions of tweets, spanning weeks, for instance, search for a hot topic. Handling these many tweets for important contents would be an unimaginable, meanwhile it is completely enormous amount of noise and redundancy that user may come across these Situation even though if filtering is allowed but its annoying.

The community of users live tweeting about a given event generates enormous contents describing sub-events that occur during an event (e.g., goals, movies, news topics etc.,). All those users share valuable information providing live coverage of events. However, this overwhelming amount of information makes difficult for the user: (i) to follow the full stream while ending out about new sub events that are occurring, and (ii) to retrieve from Twitter the main, summarized information about which are the key things happening at the event.

1.1. Clustering of data stream

The tweet stream clustering module maintains stream, and has capacity to efficiently cluster the tweets and maintain compact cluster information a scalable clustering framework which selectively stores important contents of the data, and compresses or deletes other portions. Cluster Stream is one of the most classic stream clustering methods. It consists of an online micro-clustering component and an offline macro-clustering component. number of Web services such as news filtering, text crawling, and topic detecting etc have posed requirements for text stream clustering Stream to generate duration based clustering results for text and categorical data streams. This algorithm relies on an online phase to generate a large number of micro-clusters and an offline phase to re-cluster everything our tweet stream clustering algorithm is an online procedure without extra offline clustering. We adapted the online clustering phase by incorporating the new structure TCV, and restricting the number of clusters to guarantee efficiency and the quality of TCVs.

II. TIMELINE EVENT SUMMARIZATION

We discard real-time event summarization as the activity which is providing new information about an event every time a relevant sub-event occurs. We remove a two-step process that enables to report information about new sub-events in every language. The first step is to identify at all times whether or not a specific sub-event occurred in the last few seconds or not. Next step is to choose a header tweet that describes the sub-event in the language preferred by the user.

2.1 Initialization of tweets

Initially, we collect a tweets clustering algorithm in a small number to create the initial clusters. The corresponding TCVs are initialized according to os. Next, the stream clustering process starts to incrementally update the TCVs whenever anew tweet arrives as entered by the user.

2.2 Increase Clusters

Suppose a tweet t arrives at time (ts), and there are N non passive clusters at that time. The key problem is to decide whether to attract into one of the in progress clusters or advance t as a cluster. We first find the cluster whose centroid is the closest to t. Next we get the centroid of each cluster based on formulations done above, compute its cosine similarity to t, and find the cluster Cp with the largest similar tweet.

2.3 Eleminatng unused Clusters

For most events (such as news, football matches, entertainment and new offers) in tweet streams, managing of time is important since it is not permanent for a longtime. Thus it is safe to discard the clusters representing these sub-topics when they are commonly unused. To find out such clusters, way is to estimate the average arrival time. whatever storing p percent of tweets for every increase memory costs, especially when clusters grow huge data. We employ an approximate method to get Avg p.

2.4 Merging Clusters

If the number of clusters keeps increasing then, we have an upper limit as Nmax for number of clusters. When the limit is reached its threshold, a merging process starts. The process merges cluster in a greedy algorithm. Basically, it sort all cluster pairs by their centroid similarities in a decrementing order. Starting with the most familiar pair, it merges two clusters. When many clusters are unique clusters which have not been merged with other clusters, they are merged into a new composite cluster defined by the users.

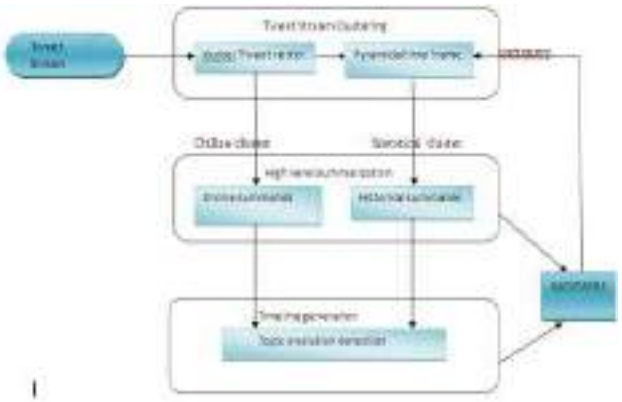


Fig: Architecture of tweet stream

Summarization at the High-level

This approach is divided into two types of summaries: 1. online and 2 .historical summaries. An online summary describes currently discussed among the users. So, the input for generating online summaries is taken directly from the current clusters handled in memory. Meanwhile, historical summary helps people understand the main happenings during a specific period, means we need to remove the influence of tweet contents from the outside of that period. This helps retrieval of the required data and even more complicated. For example the length of a user accessed time duration is H, and ending timestamp of the duration is tse.

Timeline Detection

The high need for analyzing huge contents in social medias fuels the improvement in visualization techniques. Timeline is one technique which can make analysis tasks better and quick as in presented a timeline-based hidden channel for conversations around events. This proposed the technique called ETS known as evolutionary timeline summarization to compute evolution timelines similar, which consists of a series of time-stamped summaries.

Summary-Based Variation

Tweets flow in the stream, online summaries are produced continuously by utilizing online cluster statistics in TCVs. This allows for formation of a real-time timeline. When an obvious variation occurs in the main contents in tweets (summary form), we can get a change of sub-event (i.e., a time node on the timeline. To gross the variation, we use the divergence to measure the distance between two word partitions in two successive summaries Sc and Sp (Sc is the partition of the current summary and Sp is that of the previous one)

Volume-Based Variation

Though the summary-based variation can reflect sub-topic changes. Many tweets are related to users day to day life, a sub-topic change detected from tweets description may not be significant much. At this point we consider the use of rapid raise (or spikes) in the volume of tweets in timeline, which is a common technique in present online event detection systems. we develop a spike finding method. The input, the binning process in Algorithm needs to count the tweet arrival volume in each time unit.

III. PROPOSED WORK

We propose a regular tweet stream summarization framework, defined as Sumblr, to generate timelines and summaries in the stream of tweets. We design a data structure for stream processing called TCV, and propose an algorithm called TCV-Rank algorithm for 2 types of summarization method such as online and historical summarization. Then we propose an algorithm called TCV (topic evolution detection algorithm) which produces timelines by checking three kinds of variations. Regular testing on real Twitter data sets demonstrate the efficiency and effectiveness of our framework for the requirements of user.

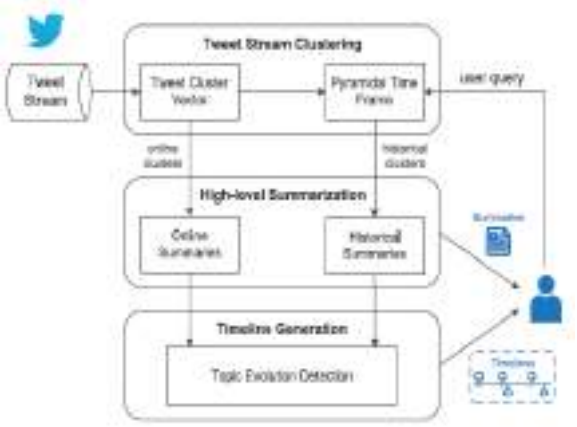


Fig.: The Framework of Sumblr

Sumblr is abbreviated as continuouS sUMmarization By stream cLusteRing. The framework consists of three major division,1. Tweet Stream Clustering module, 2. High-level Summarization module and 3.Timeline Generation module. In tweet stream clustering module, we design an efficient tweet stream clustering algorithm, for allowing the effective clustering of tweets with only one pass over the data we use an online algorithm allowing. This algorithm has two data structures to keep important tweet information in clusters. The initial one is a novel compressed structure known as tweet cluster vector TCV. TCVs are analyzed as potential sub-topic delegates and secures dynamically in memory during stream processing. The another structure is the pyramidal time frame narrated as PTF, which is used to save and organize cluster snapshots at variety of moments, thus allowing historical tweet data to be retrieved by any arbitrary time durations. This module supports generation of two kinds of summaries:1. online and 2. historical summaries. (1)we propose a TCV-Rank summarization algorithm by allude to the current clusters maintained in memory, to generate online summaries. This high-level summarization algorithm begins its computation by centrality scores for tweets kept in TCVs, and selects the top-ranked ones in terms of novelty content and coverage. (2) To compute a historical summary where the user specifies an haphazard time duration, we very first retrieve two historical cluster snapshots from the PTF with respect to the two endpoints one is the beginning and other one is ending points of the duration. Then, the TCV-Rank summarization algorithm is used to generate summaries, based on the difference between the two cluster snapshots. The elemental of the timeline generation module is a topic evolution detection algorithm, which utilize online or historical summaries to produce real-time or range timelines.

The algorithm supervises quantified variation during the course of stream processing. In addition of a new node on the timeline a enormous fluctuation at a particular moment implies a sub-topic change. In our design, we consider 3 different factors in the algorithm. First, we mark variation in the main contents discussed in tweets, in the form of summary. To calibrate the précis based variation (PUM), To measure the distance between two word distributions in two serial summaries, we use the Jensen-Shannon divergence (JSD). Second, we monitor the volume-based variation (VOL) which reflects the importance of sub-topic changes, to find high level increases (or "spikes") in the volume of tweets over time. Third, we state the sum-vol variation (SV) by joining both effects of summary content and significance, and detect topic evolution whenever there is a break in the consolidate variation.

FIRST STEP: SUB-EVENT DETECTION

The first part of the event summarization system corresponds to the sub-event detection. The system has to check at all times whether or not a relevant sub-event has occurred, irrespective of how the stream will continue to evolve. Before the start of an event, the system is provided with the time that it begins, as scheduled in earlier event, so the system knows when to start looking for new sub events. With the target of developing a real-time sub event detection method, users depend on the fact that relevant sub-events trigger a massive tweeting activity of the community. The more important a sub-event is, the more users will tweet instantly about it almost immediately. This is rejected as peaks in the histogram of tweeting rates. In the process of identifying sub-events, we aim to compare 2 different ideas: (i) only sudden increase with respect to the recent tweeting activity, and (ii)By considering also all the activity which is seen previously during a game, so that the system acquires from the modification of the viewers. We compare the following two methods that relay on these 2 ideas:

1.Increase: this increase approach was introduced by Zhao et al. It considers that an important sub-event will be reacted as a sudden raise in the tweeting rate. For time periods defined at seconds 10, 20, 30 and 60, this method checks previous time frame for any of those history if the rate of tweeting increases by at least 1.7 from the previous time. If the expansion actually occurred, it is considered that a sub-event occurred. This method is that not only outstanding tweeting rates would be submitted as sub-events, but also lowers the rates that are foregoing by even lower rates which is major drawback of it.

2. Outliers: This introduce approach relies on whether the given timeframe stands out from the regular tweeting rate seen so far during the event for a tweeting rate (not only from the previous time frame). We set the time period in seconds 60for this approach. Initial 15 minutes before the game starts, the system begins to learn from the tweeting rates, to and out what is the approximate audience of the event. When the start time approaches, very first the system begins with the detection process of sub-event. The system considers that a sub-event occurred when the tweeting rate represents and activity seen before an outlier is the one compared both of them. If it is (tweeting rate) above 90% of all the earlier visualized tweeting rates, the current time frame will be reported as a sub-event. This threshold has been set a priority without optimization. By comparing the current tweeting rate to all the rates seen before, the outlier-based method step-by-step learns while the game advances. Different from the increase-based approach, our method considers the specific viewer of an event which is an advantages of it, and that consecutive sub-events can also be finds if the tweeting rate remains uniform without increase. Accordingly, this method will not consider that a sub-event occurred for low tweeting rates foregoing by even lower rates, as opposed to the increase-based approach. time elapsed tweet rate
500
1000
1500
2000
2500
soccer game (Argentina vs Uruguay), where several can be seen.
Since the notations are limited to minutes on the references, we round down the outputs of the systems to match the

reference. Also, the timestamps noted for the reference are not entirely shortened, and therefore we accept as a correct guess an automatic detection of sub-events that varies by at most 1 minute from the reference. This evaluation method enables us to compare the two systems to conclusion which of them performs best. For improving effectiveness of summarization system in order to provide a brief and accurate summary, is achieved by keeping the number of sub-events small. The out performance of this(outlier-based) approach shows the significance of taking into account the audience of a specific game, as well as the effectiveness of learning from previous activity throughout is seen.

SECOND STEP: TWEET SELECTION

The last part of the summarization system is the tweet stream. Only when first step reports that new sub-event has occurred then only next step is activated .Once the system has determined that a sub-event occurred, the selector is provided with the tweets corresponding to the time (in minute) of the sub-event. From those tweets, the system has to choose one as a head of the tweet that tells what has happened. This tweet must provide the main information about the sub-event, so the user will get to know what happened and can track the event. Here we compare two tweet selection methods, depending only on information stored within that minute of the sub-event. We use outlier-based sub-event detection approach to test them on the output described above, as the approach with best performance for the rest step. we get a ranking of all the tweets, to select a representative tweet. To do so, we earn each tweet with the average of the values of the terms that it contains. The more head clusters are the terms contained in a tweet, the more representative will be the tweet itself. To define the values of the terms, we tally 2 methods: (i) only the sub-parts can consider tweets(to give highest values to terms that are used frequently within the sub-event), and (ii) taking into the user account also the tweets are sent before throughout the game, so that the system can make a fluctuations from what has been the very common vocabulary during the event (to give highest values to terms that are particularly used within the time(min) and not so periodically earlier during the event). We were using these following famous approaches to implement these two ideas:

1. Tweet Frequency (TF): each term is given the value of its frequency as the number of appearance within the minute, nevertheless of its prior use.
2. KLD: know as Kullback-Leibler divergence we use this to measure how frequent of term t within the sub-event (H),During the game until the previous minute(G) it also considering how frequent it is. Thus, KLD will give a higher weight to terms frequent within the minute that were low repeated frequency during the game analysis. Rather provide higher rates to specific terms within the sub-event. In all along the game this may allow to get rid of the common vocabulary,

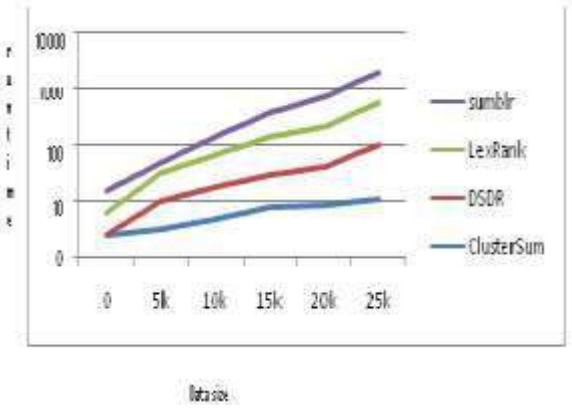
DKL (HkG) = H(t) logH(t)G(t) called as equation1

With these two approaches, the sum of values for each terms contained in each tweet results in a strength for each tweet. With weights given to all tweets, tweets are sent during the sub-event are considered and creates ranking of it , where the tweet with highest weight ranks list.

Baseline methods:

In Existing summarization methods have not been designed to handle continuous summarization. In this way, here implement the sliding window version of the above three algorithms, namely Cluster Sum, LexRank, and DSDR.

In our experiments, we find similar trends in the comparison of precision, recall and F-score be tween



opposed a prototype called Sumblr which supported continuous tweet stream summarization. To compress tweets into TCVs and maintains them in an online fashion we use sumblr clustering algorithm. Then, for to generating online summaries and historical summaries with arbitrary time durations it make use of TVC-Rank summarization algorithm. Automatic topic evolution done by allowing Sumblr to produce dynamic timelines for tweet streams. The experimental results make obvious the competence and success of our method.

CONCLUSION

We proposed a new technique for continuous tweet stream summarization called Sumblr. this employs a clustering algorithm to reduce tweets into TCVs and maintains them in an online fashion. Then, for to generating online summaries and historical summaries with arbitrary time durations it uses a TCV-Rank summarization algorithm. By allowing sumblr to produce dynamic timeline for streams of tweet, automatic topic evolution can be detected. The experimental results exhibit the capacity and productiveness of our idea.

For future work, we aim to develop a multi-topic version of Sumblr in a distributed system, and analyses it on more complete and huge data sets. We design to develop a multi topic version of Sumblr in a spread system, we can also try to do estimation on more complete and large-scale datasets.

ACKNOWLEDGEMENTS

The successful completion of any work depends upon the cooperation and help of many people and not just those who directly execute the work. It is difficult to express in words our profound sense of gratitude to those who helped us. But I make a humble attempt to do so.

I wish to place on record my sincere thanks to Head of the Department and the guide of the project computer science and engineering for the encouragement and support.
Also I thank the members of the faculty of CSE department whose suggestions helped me in the course of this Seminar

REFERENCES

[1] C. C. Aggarwal, J. Han, J. Wang, and P. S.Yu, ―A framework for clustering evolvingdata streams,‖ in Proc. 29th Int. Conf.VeryLarge Data Bases, 2003, pp. 81–92.
[2] T. Zhang, R. Ramakrishnan, and M. Livny,BIRCH: An efficient data clusteringmethod for very large databases,‖ in Proc.ACM SIGMOD Int. Conf. Manage. Data,1996, pp. 103–114.
[3] P. S. Bradley, U. M. Fayyad, and C. Reina,Scaling clustering algorithms to largedatabases,‖ in Proc. Knowl. Discovery DataMining, 1998, pp. 9–15.
[4] L. Gong, J. Zeng, and S. Zhang, Textstream clustering algorithm based onadaptive feature selection,‖ Expert Syst.Appl., vol. 38, no. 3, pp. 1393–1399, 2011.
[5] Q. He, K. Chang, E.-P. Lim, and J. Zhang,Bursty feature representation for clusteringtext streams,‖ in Proc. SIAM Int. Conf. DataMining, 2007, pp. 491–496.
[6] J. Zhang, Z. Ghahramani, and Y. Yang, Aprobabilistic model for online documentclustering with application to noveltydetection,‖ in Proc. Adv. Neural Inf.Process. Syst., 2004, pp. 1617–1624.
[7] S. Zhong, Efficient streaming textclustering,‖ Neural Netw., vol. 18, nos. 5/6,pp. 790–798, 2005.
[8] C. C. Aggarwal and P. S. Yu, On clusteringmassive text and categorical data streams,‖Knowl. Inf. Syst., vol. 24, no. 2, pp. 171–196, 2010.
[9] R. Barzilay and M. Elhadad, Using lexicalchains for text summarization,‖ in Proc.ACL Workshop Intell. Scalable Text
Summarization, 1997, pp. 10–17.
[10] W.-T. Yih, J. Goodman, L. Vanderwende,and H. Suzuki, Multidocumentsummarization by maximizing informativecontentwords,‖ in Proc. 20th ntJoint Conf.Artif. Intell., 2007, pp. 1776–17

Overview of Big Data Visualization

Sagargouda S.Patil
Department of Computer Science and Engineering, PES University, Bangalore.
Email-id:sagar.cs.kle@gmail.com

ABSTRACT- Day by day, data is growing at a faster rate in every field. Data which are very large in size and cannot be processed using traditional database systems are known as big data. Before the days of data visualization and big data, businesses struggled to grasp concepts such as customer behavior and fraud. Big data helped these insights come to light. Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner and can be experimented with different scenarios by making slight adjustments. This paper discusses about big data Visualization, its challenges and tools.

Keywords— big data, big data visualization, data visualization, interactive visualization, visualization tools.

I INTRODUCTION

Big data [1] Visualization is a representation of data in graphical format or pictorial format. It enables us to evaluate things visually so that we can identify new patterns. There are some key points to be considered for big data visualization such as: (1) Meta data (2) Visualization tools must be interactive. There are some myths related to visualization are as follows:

- Every data must be visualized
- Good data should be visualized
- Visualization will be always evident right decision
- Visualization leads to the inevitability

Big data visualization is not easy as compared to traditional data sets. In larger data set visualization, many researchers use feature extraction and modeling techniques to reduce data size.The Following table shows the benefits of big data visualization according to a survey report [1].

Benefits	Percentages (%)
Improved decision-making	77
Better ad-hoc data analysis	43
Improved collaboration/information sharing	41
Provide self-service capabilities to end users	36
Increased return on investment (ROI)	34
Time savings	20
Reduced burden on IT	15

Table 1. Benefits of visualization tools

II LITERATURE SURVEY

This section focuses on the conventional visualization methods and how big data visualization came into the picture.
Conventional visualization method [2] is a traditional method which uses tables, bar chart, pie chart, histogram, Venn diagram, flow chart, entity relationship diagram for data visualization. These methods do not work well for huge datasets containing structured and unstructured data respectively.Given below are the 2 additional methods which are used often in conventional visualization.

(1) **Parallel Co-ordinates**: It is used to plot data elements across multi-dimensions.

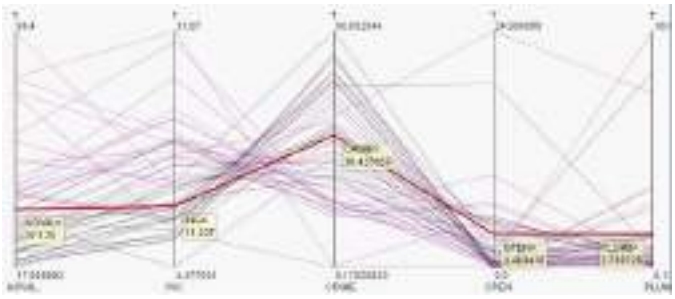


Fig 1. Parallel Coordinates

(2) **Semantic network**: It is a logical relationship between different concepts.
Conventional data visualization conveys that these visualizations could use in business analytics and other data systems, but cannot be used for the scientific process.

Visualizations are not only static, but also they can be interactive. Interactive visualization [4] help in zooming (zoom in or zoom out) ,focus, overview and detail.
Below given are the Steps for interactive visualization.
1] Selecting
2] Linking
3] Filtering
4] Remapping

A. **Fig 2. Interactive brushing and linking**

Big data visualization is important because of the following reasons.

- It provides clear knowledge about patterns of data
- Detects hidden structures in data
- Identify areas that need to be improved
- Help us to understand which products to place where
- Clarify factors which influence human behaviour

III. CHALLENGES OF BIG DATA VISUALIZATION

This section describes the challenges of big data visualization [3][10].Visualization based methods of big data challenges are presented by "**four Vs**" as follows:

- **Volume**= the methods are developed to work with a large number of datasets.
- **Variety**= the methods developed to combine as many data sources (structured +unstructured data).
- **Velocity**= with the methods, business can replace batch processing with real time processing.

Visualization of heterogeneous big data (structured, unstructured, semi-structured) is a biggest problem. Because of the big data size, parallelization is a challenge in visualization and Speed is a desired factor for big data analytics. So, effective data visualization is a key point of the discovery process in the field of big data [3].

Following are some of the problems of big data visualization.

- **Visual noise**: Objects which are present in the data set are related to each other. So users cannot divide them into separate objects on screen.
- **Information loss**: There is information loss while reducing data sets.
- **Large image perception**: Visualization methods are not only related to aspect ratio and resolution but also it takes physical perception into consideration.

Potential solutions to some challenges and problems of data visualization are [6][7]:

- **Meeting the need for speed**: One solution is to deal with "Hardware" by increasing memory and massive parallel processing can be used.
- **Understanding the data**:"Proper domain Expertise" is solution.
- **Displaying meaningful results**: solution is to cluster data into smaller groups that are visible effectively.
- **Dealing with outliers**: solution is to remove outliers from data or create separate chart for outliers.

IV. BIG DATA VISUALIZATION TOOLS

This section discusses about different big data visualization tools most of which run on the Hadoop platform [3][10] having following modules:

- Hadoop Distributed File System(HDFS)
- Hadoop Map Reduce
- Hadoop YARN(Yet Another Resource Negotiator).

Some of the big data visualization tools available are as follows:

- **Statistical Analysis System(SAS) Visual analytics**: It uses intelligent auto charting to create best possible visual data we select.
- **Polymaps**: It is a free java script library for making interactive maps in modern web browsers.
- **Flot**: It is a java script plotting for JQuery. It supports visualization for interactive charts, data points, zooming and stacked charts.
- **Google Maps**: This tool enables developers to build visual mapping programs for any website or application.
- **Microsoft Excel**: It is also platform for data analysis and creates visualization and represents data in the form of charts and graphs.
- **R-Project**: 'R' is a free software environment for statistical and graphics. It helps in data manipulation, calculation and analysis.

- **Tableau Public**: It is a simple and user-friendly tool to create an iterative data visualization quickly and puts them on the website [5][9].

CONCLUSION

The strong and growing trend toward visualization-based data discovery tools is expected to continue in the years to come, as more businesses seek better, more cost-effective ways to derive meaning from their big data. Big data Visualization provides an immense opportunity not only to manage the growing volume, variety, and velocity of new and existing data but also to turn that data into value. It allows access to challenging data sets, it allows exploration, can be fun, and provides useful information in an efficient way. This paper discusses big data visualization basics, tools, challenges which will benefit future researchers of big data visualization to continue their work based on the concepts discussed here.

ACKNOWLEDGEMENT

I am grateful to the valuable guidance provided by Prof. Nitin V Pujari, HOD, Department of Computer Science& Engg, PES University, Bangalore and Dr. Usha Sakthivel, HOD, Department of Computer Science& Engg,RRCE, Bangalore and special thanks to Prof. Rashmi G, Department of Computer Science& Engg, Nandi Institute of Technology and Management Sciences, Bangalore, Prof. V.R.Badri Prasad, Dept of Computer Science & Engg, PESIT, Bangalore and Prof. Dinesh Singh, Dept of Computer Science & Engg, PESIT, Bangalore for providing me a valuable help and support.

REFERENCES

[1] M. Khan, S.S. Khan, Data and Information Visualization Methods and Interactive Mechanisms: A Survey,*International Journal of Computer Applications*, 34(1), 2011, pp. 1-14.

[2] Intel IT Center, Big Data Visualization: Turning Big Data Into Big Insights, White Paper, March 2013, pp.1-14.

[3] V. Sucharitha, S.R. Subash and P. Prakash , Visualization of Big Data: Its Tools and Challenges,*International Journal of Applied Engineering Research*, 9(18), 2014, pp. 5277-5290.

[4] P. Simon, The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions, *Harvard Business Review*, June 13, 2014, pp. 1-8.

[5] C.L. P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, *Information Sciences*, 275 (10), August 2014, pp. 314-347.

[6] B. Porter, Visualizing Big Data in Drupal: Using Data Visualizations to Drive Knowledge Discovery, Report, University of Washington, October 2012, pp. 1-38.

[7] T. A. Keahey, Using visualization to understand big data, Technical Report, IBM Corporation, 2013, pp. 1-16.

[8] P. Fox and J. Hendler, Changing the Equation on Scientific Data Visualization, *Science*, 331(11), February 2011, pp. 705-708.

[9] I. B. Otjacques, UniGR Workshop: Big Data- The challenge of visualizing big data, Report, Gabriel Lippmann, 2013, pp. 1-24.

[10] H. Childs, B. Geveci, J. Meredith, K. Moreland, C. Sewell, E.W. Bethel, T. Kuhlen, W. Schroeder, Research Challenges for Visualization Software, Joint Research Report of Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Sandia National Laboratories, Los Alamos National Laboratory, RWTH Aachen University (Germany), May 2013, pp. 1-11.

Predictive, Corrective actions of Energy management and benefits through Data Analytics with Intelligent Reporting

Spurthi. R .Terwai#1, Shamshekhar. S. Patil #2

Dept. of Computer Science and Engineering,
Dr. Ambedkar Institute Of Technology, Bengaluru, India.
#1spurthiterwai@gmail.com, #2shamshekhar.patil@gmail.com

Abstract- This paper presents the software model of energy management adapting standard ISO 50001 and the upgraded version of its model with several improvisations to the same. The application of the sustainable energy management system helps to increase the efficiency in energy management aims at reducing production of energy costs and improving efficiency of acquiring energy parameters from the Energy meters. The paper also analyzes the efficient implementation of energy management with dynamic reporting facility for the required Energy parameters like the Energy consumption ,current ,voltage and many more comparisons involving cost calculation , relative energy utilization and trend prediction. The Energy parameter values are efficiently stored in the Database with priority scheduling of highest polled quantum value .The paper also discusses the advantage and disadvantage of implementing ISO 50001. Survey results and use cases show the suitability that the model can be deployed immediately and also propose a roadmap of implementing the same with cloud server storage.

Keywords- ISO50001, Modbus, Energy Management, sense and store, dynamic reporting.

I.INTRODUCTION

Energy conservation refers to reducing energy consumption through using less of an energy service. Energy Management and conservation plays a prime role in the modern world of Globalization. Optimization and Management of energy consumption is critical considering economic and ecological implications. Detailed analytics involving trend prediction are applied and a visualization engine renders graphs and alerts through the visualization feature of the software. By quantitatively and qualitatively assimilating energy consumption behavior, not only are device-level optimization techniques proposed, but rather an intuitive way to consciously and judiciously consume energy is implied. Some of the ISO 500001 standards are also applied.

Energy Engineering – The Need of the Hour

It is predicted that Global energy needs are set to increase 40% by 2030. This would bring in immense pressure on utilization of existing resources and thus prices. Energy management systems are becoming very critical in today's context due to increased consumption and decreased resource availability. The IOT Cloud (BIC) service provides a tailor made and effortless platform for seamless integration of physical node data into the internet. For instance, a day to day display of per capita energy usage could go a long way in creating awareness that a single watt saved multiplied by a lot of people can end up saving a lot of kilowatts of energy. **Sustainable architecture** minimizes the negative environmental impact of buildings by efficiency and moderation in the use of materials, energy, and development space[1]. Energy Management Systems (EMnS) are integrated, computerized systems for monitoring and controlling energy-related building services. Small buildings may have independent controls for energy systems, but for larger or more complex buildings, sites with a number of buildings, or organizations with buildings on a number of sites, the integration of systems operation through a single Building Energy Management System offers greater control and can achieve significant savings [8]. Increasingly, simple EMnS is being developed that are suitable for even the smallest buildings.

II. EASE OF USE

The purpose of this paper is to establish a standard method for monitoring and reporting on the Energy performance of commercial buildings. It determines the energy consumption, electrical Energy demand, and on-site energy production in existing commercial buildings of all types. It mainly provides a software perspective for improving efficiency of Energy management in commercial buildings. To provide this ISO 50001 standards are followed. Also, a comparative study on Energy parameters with and without following the standard has been done.

III. GOALS OF ANALYSIS.

1. To design and implement an embedded system for energy consumption monitoring and control at commercial buildings.

2. The solution has currently been deployed to monitor the complete campus, providing valuable information with visualization and analysis to the facilities team.

3. Energy Management mainly includes the following stages,

- Data Acquisition
- Data Processing
- Data Analytics
- Data Visualization
- Control/feedback/Optimization

4. The project aims to design and implement an embedded system for energy consumption monitoring and control at any commercial building.

5. The solution has currently been deployed to monitor the complete campus, providing valuable information with visualization and analysis to the facilities team.

A comparative study on Energy metrics with and without ISO standards is done and few of the Energy parameters are compared and analyzed. Also, big data analytics and web services might be used to improve the efficiency of data storage that is Energy data can be stored on priority basis (highest polled meter or the main meter) on the cloud storage space.

IV. LITERATURE REVIEW

The Energy Management Solution (EMnS) helps answer these vital questions in both qualitative and quantitative manner providing vital data points for the end user and enabling him take decisions that make energy management more efficient EMnS aims to build a technological platform for energy management in buildings, offices and commercial spaces.

The ongoing revolution of everyday objects wirelessly connected to the network, called the Internet of Things (IoT), is creating interesting and unexpected opportunities in reducing energy consumption and improving environmental comfort in buildings.

A. What is ISO50001?

ISO50001 is a standard designed to manage energy across the entire international commercial sector, affecting energy use, which can be monitored and influenced by an organization. The purpose of ISO50001 Energy Management System (EMnS) standard is to enable organizations to establish the systems and processes necessary to improve energy performance, including energy efficiency, use and consumption[2]. The implementation model also applies Plan-Do- Check-Act (PDCA) cycle which consists of five clauses i.e. Energy Policy, Planning, Implementation, Monitoring and Corrective Action and Management Review [9].

B. Why ISO50001?

An ISO50001 energy management system (EMnS) is applicable to all the activities under the organizational control. This would enable organization to achieve its policy commitments, take action as needed to improve its energy performance and demonstrate the conformity of the system to the requirements of the International Standard [2]. It is also an instrument to quantify the energy use and consumption, and to plan energy efficient practices by fine tuning the operational controls and organizational behavior. Optimal energy performance should be achieved using existing resources, with minimum or low cost investment, through the housekeeping efforts prior to suggesting improvement that require financial investments [3].

C. Challenges in ISO50001 implementation

Despite valuable benefits, achieving and sustaining the EMnS and energy efficiency are challenging for organizations. Survey mainly records three major challenges i.e. lack of management commitment [2], lack of communication and understanding at all levels [10], and design error due to not prioritizing energy efficiency during the design stage [8]. In addition to these, particularly in Asia, lack of financial support and lack of policies by the government are also barriers that need to be tackled [1]. Therefore, for the organizations to enjoy the benefits of EMnS implementation, it is essential to overcome these challenges.

D. Challenges faced in EMS implementation

The Energy meters are generally functional. Sometimes, the values can become null or zero as they might become non-functional or reached threshold values .So, measures have to be taken to handle these exceptions both in database storage and software development. Despite many available connecting techniques, serial communication is used and therefore there is deferred calculation of Energy parameters even though it is addressed as current data .This depends on the logging interval.

V. METHODOLOGY

Metric Definitions.

Functional Area: is measured as the Gross Interior Floor Area of all spaces that have energy use included in the Building Energy Use metric. This may include the whole building, or it may be divided into separate areas with different functions such as office, mercantile, lodging, and parking [5].

- Building Electrical Demand (kW or kVA): Peak electrical demand of the electrical building Energy Use during the month.
- Building Electrical Demand Intensity (kW/ft2, kW/m2, kVA/ft2, or kVA/m2)= Building Electrical Demand ÷ Functional Area
- Building Energy Use Intensity (BEUI)(kWh/ft2 kWh/m2, Btu/ft2, or Joules/m2)= Building Energy Use ÷ Functional Area
- Building Lighting Energy Use(kWh, Btu, or Joules)= Installed Lighting Energy Use + Plug-in Lighting Energy Use + Facade Lighting Energy Use Reported as: Monthly totals (tabular), monthly daily averages (graphical), annual total, graph of peak day in each billing period .
- Building Purchased Energy Cost Intensity(CurrencyYear/ft2 or CurrencyYear/m2) = Building Purchased Energy Cost ÷ Functional Area Reported as: annual value

For cost, calculations the energy utilization is multiplied with the unit rates separately for Peak hour, Industrial hour and Night hour with the respective unit rates depending on the requirement ,that is , hourly, per day, per week ,per month and annually.

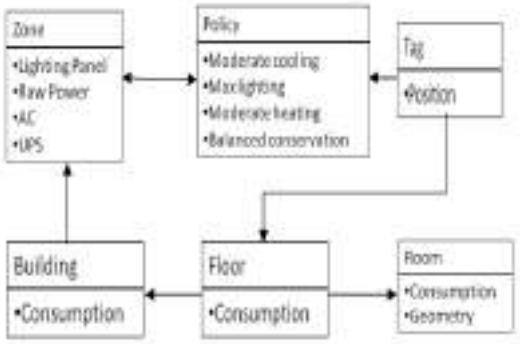


Fig1.Building model in UML

VI. SYSTEM ARCHITECTURE .

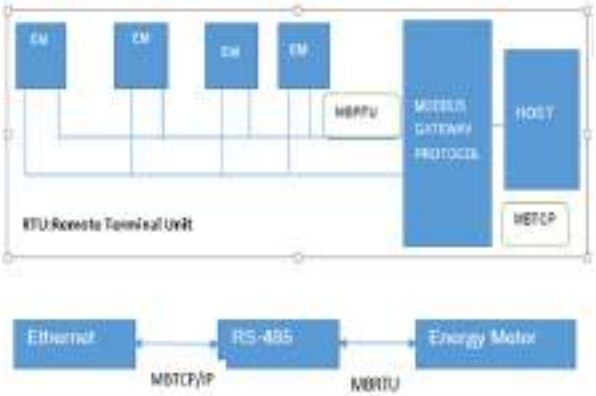


Fig 2. System architecture

This study uses the data collected from Energy meters as input to the software. Energy meters are connected using MOBBUS RTU to the Modbus gateway protocol. This gateway is connected to host by MBTCP[10]. It is connected in round robin fashion with a given time quantum to fetch Energy data from each Energy meter. The highest value read along with all other values of Energy parameters is recorded and for the second round, the highest valued meter is polled with priority. So, small modification is made to round robin with priority scheduling of Energy meter polling pattern.

Dynamic reporting feature is also developed where costs, energy utilization details with respect to each meter and zones is available .The reports are created on selecting the required date and can be viewed either in the Report Viewer or as PDF. It also includes the implementation of ISO50001 EMnS as described in the "Introduction" section. Deviation from standard value is calculated for ISO implementation .To evident the implementation activities, the communications, analysis and results were recorded in EMnS tool. The EMnS Tool is an Excel workbook sending and storing data to database via TCP/IP port and also on cloud with SAAS service and storage service provided by AWS or an IoT private cloud provided by the company or customer.

WORK FLOW

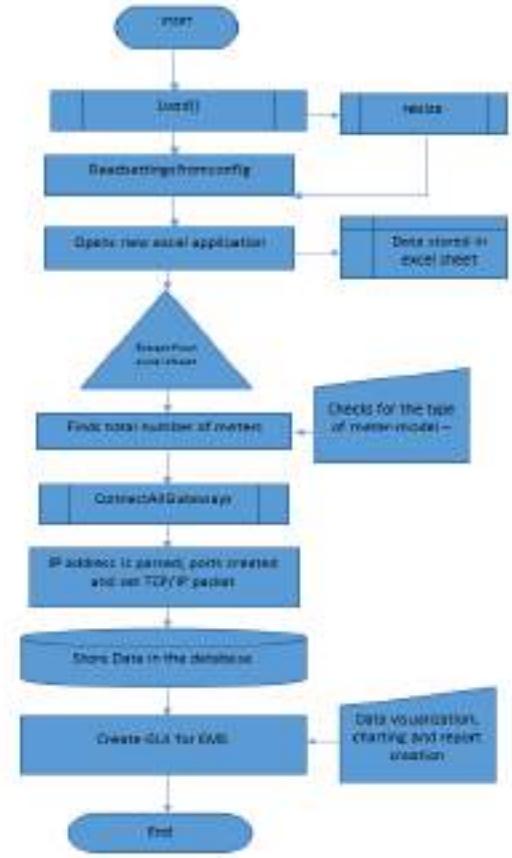


Fig 3. Flow of work

The data acquisition process begins with reading settings from the configuration file which has the list of IP addresses of Energy meters that are treated as hosts. Also Device ID's are assigned to each Host .Once all meters are read with the meter model , all gateways are connected , here the MODBUS gateway protocol is used .IP address are parsed and data is parsed and formatted accordingly . TCP/IP packets are used for sending and receiving through the TCP/IP layer of Network .The data is formatted with respect to the type of Networking Model followed and stored in the database. This data is used for visualization and analysis. Also, reporting with ISO 50001 standards is adapted to increase efficiency of energy utilization. Modbus TCP/IP and Modbus RTU are used for communication between Energy Meters and Modbus Gateway Protocol and the same and Host systems respectively. Results of the EMnS implementation are shown in the next section.

VIII. RESULTS

Results mainly include the visualization and reporting of Energy parameters.

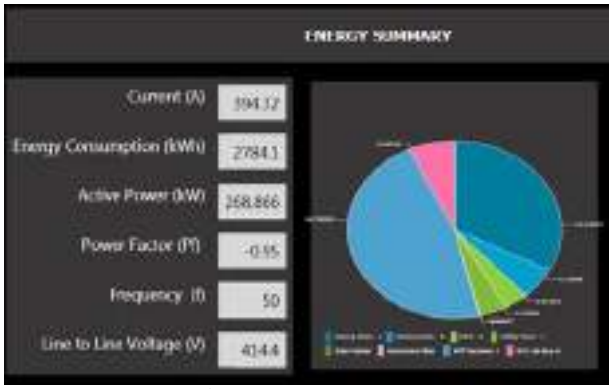


Fig 4. part of Dashboard representing energy params

The above graph represents the Lighting Panel Zonal division. It displays the meters included in the Lighting Panel zone and few of the Energy parameters like Current, Energy Consumption, Active power, Power factor, Frequency, Line to Line voltage around 10 am from 12.01 a.m.

TABLE 1. COST TABLE

	Energy Meters	Cost
Cost Previous Day	Vital Loads	73334
	Main Incomer Load	159461
Cost Previous Week	Vital Loads	928711
	Main Incomer Load	1279468
Cost per month till date	Vital Loads	1126768
	Main Incomer Load	2199131

The table shows the cost comparisons of the vital loads and the main incomer load with data analyzed per day, per week and month. Also, the required date can be selected from the calendar option and the particular day's Energy consumption and other parameters can be retrieved.

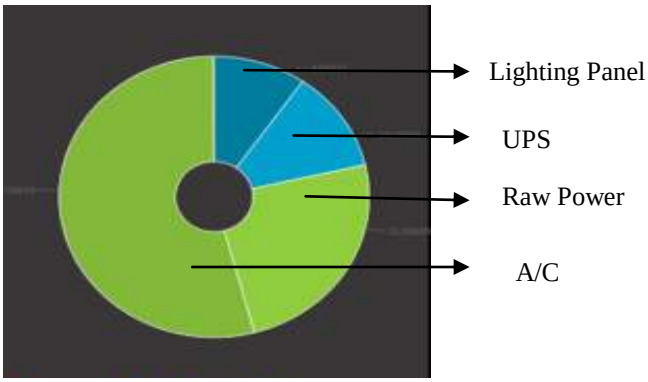


Fig 5. Zonal distribution

Pie chart shows the zonal energy distribution between
1) Lighting Panel.2)UPS3).Raw Power 4).A/C

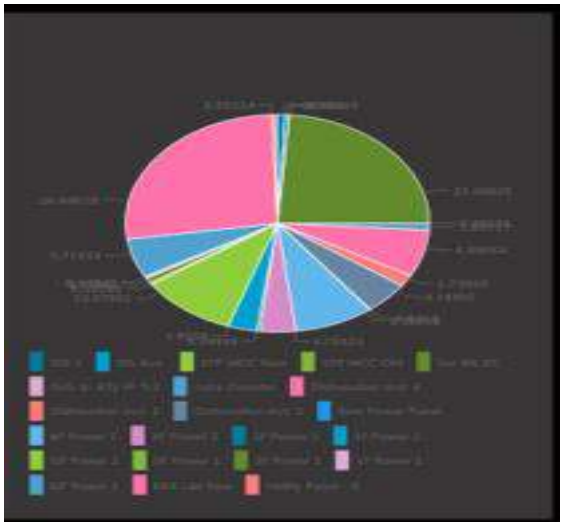


Fig 6. Chart depicting various meters in Zones

The pie chart shows the Energy Meters per zone and also the Energy consumption value .Here Energy meters grouped under Raw Power is shown which includes many meters like MCC mains, 2F Power and many more.

Fig 7.Emoticon

Emoticons are to be added to measure relative energy consumption- the energy consumption value is compared to the previous day's value at the given time .If the current days' consumptions is less than the previous one , then it is indicated by green color else red .
Green smiley indicating efficient utilization of Energy.

A. Advantages of the proposed approach

- EMS is intended to provide organizations with a recognized framework for integrating energy performance into their management practices.
- Multinational organizations will have access to a single, harmonized standard for implementation across the organization with a consistent methodology for identifying and implementing improvements.
- Cost estimation
- Outlier frequency analysis.
- Design of Algorithms for Data Analytics and Optimization
- Data can be better analyzed and trend predictions and further improvements can be performed.
- Dynamic reporting is added for analyzing making the Energy data readable.
- A Dashboard is to be generated with the display of Energy parameters. The dashboard should also

include the Graphical Representation of Energy parameters described.

The real-time visualization and notification system has empowered the facilities team to monitor and optimize energy at the campus. The little savings gained are first steps towards a greater milestone for the planet.

CONCLUSION

In this paper, an optimized method for classification of building energy-consumption data has been proposed. First, the features of the daily energy-consumption profiles are extracted by using MODBUS protocol connected serially. Data analysis and data visualization techniques have been used to extract information in a customer adaptable manner. Trend analysis is done based on historical data. As a result, abnormal energy consumption can be identified in real time, enabling building managers to investigate and correct problems as they occur. Also, dynamic reporting feature is made available where current data can be represented and analyzed for further improvisation of data acquisition techniques.ISO 50001 standards have been followed and also a comparative study with and without In this system, the procedure followed are computationally efficient and robust, therefore, can feasibly be integrated into existing building energy management systems. Further work will aim to build on this classification technique to provide additional tools for automated analysis of metered building and optimizing the storage capacity.

ACKNOWLEDGMENT

A special thanks especially to Dr.Ambedkar Institute of Technology for providing the assistance and Computer Science Department for the Guidance; Mr. Shamshekhar .S.Patil for being part of the team.

REFERENCES

[1] Intelligent Energy Management System: Digitizing Energy *[https://www.gedigitalenergy.com/uos/ems.htm]*

[2] ISO 50001 Energy Management [http://www.iso.org/iso/home/standards/management-standards/iso50001.htm]

[3] Schneider Electric India [*www.schneider-electric.com/site/home/index.cfm/in]*

[4] Guide to Energy Management, Seventh Ed. 7th Ed. by Barney L. Capehart Ph.D. CEM (Author), Wayne C. Turner Ph.D. PE CEM (Author), William J. Kennedy Ph.D. PE (Author)

[5] Energy Management Handbook, 6thEdition ,Wayne C Turner and Steve Doty

[6] Xiaoli Li, Chris P. Bowers, and Thorsten Schnier "Classification of Energy Consumption in Buildings With Outlier Detection", in Proceedings of IEEE. Transactions on Industrial Electronics (Volume:57 , Issue: 11)

[7] Zheng O'Neill,Madhusudana Shashanka, Xiufeng Pang ,Prajesh Bhattacharya, Trevor Bailey, and Philip Haves United Technologies Research Center, East Hartford, USA Lawrence Berkeley National Laboratory, Berkeley, CA, USA " Real Time Model-Based Energy Diagnostics In Buildings", 2011 IEEE

[8] Han Chen, Paul Chou, SastryDuri, Hui Lei, Johnathan ReasonIBM Thomas J. Watson Research Center, 19 Skyline Dr, Hawthorne, NY 10532{chenhan, pchou, sastry, hlei "The Design and Implementation of a Smart Building Control System", reason}@us.ibm.com,IEEE

[9] Universidade Nova de Lisboa - Faculdade de Ciências e TecnologiaMonte da Caparica, Portugal"Energy Consumption Monitoring System for Large Complexes",IEEE.

[10] Modbus-IDA, www.modbus.org

[11] The Energy Management According to ISO 50001:2011 Standard and AEMAS Scheme Feasibly Implement in Viet Nam

Implementation of an Efficient MongoDB NoSQL Explorer for Big Data Visualization

Chaitanya.P[#1], Ranjan H.P[#2], Kiran T.S [#3] Anitha.K*[4]
UGstudents, DeptofCSE, RRCE, Bangalore, Karnataka, India[1,2,3]
Associate Prof. Rajarajeswari College of Engineering, Bangalore, India[4]

ABSTRACT- With the emergence of Big Data, the use of NoSQL (Not only SQL) has increased among internet companies and other enterprises. Benefits include horizontal scaling, finer control over availability and simplicity of design. NoSQL databases are considered as an alternative to relational databases, as its schema less data model is considered to be better for handling the large volumes of structured and unstructured data. This paper aims to introduce the concepts behind NoSQL and provide arguments for and against adopting NoSQL. A small library application has been developed to assess the stated benefits of NoSQL and NOSQL and compare its performance with the existing systems like MySQL/ Oracle and a re-usable tool called Mongo-sight is developed to visualize the data stored in it.

Keywords— NOSQL, SQL, Databases, structured data, unstructured data, Big Data, Mongo-Sight, MongoDB.

I. INTRODUCTION

Big data is a buzzword, catch-phrase, which means massive volume of both structured and unstructured data which is so large is difficult to process using traditional database and software techniques. It is estimated that volume of data is increasing 40% per year, and will grow 44 times between 2009 and 2020. Most of the data is unstructured as it is in the textual form. For example NoSQL databases are mostly used to collect and store data related to social media.

II. OVERVIEW OF NOSQL

NoSQL databases are non-relational and can accommodate unstructured data. It is not a replacement for SQL database but rather compliments it, both these technologies can coexist.

One of the key differences is that SQL databases have rigid schemas while NoSQL offeres a flexible schema that can be altered without service disruption.

There are four categories of NoSQL databases:

A key-value store, or key-value database, is a data storage paradigm designed for storing, retrieving, and managing associative arrays, a data structure more commonly known today as a dictionary or hash. Dictionaries contain a collection of records, objects which in turn have many different fields within them, each containing data. These records are retrieved & stored using a key that uniquely identifies the record, and is used to quickly find the data within the database.

Key-value stores work in a very different fashion from the better known relational databases (RDB). RDBs pre-define the data structure in the database as a series of tables containing fields with well defined data types. Exposing the data types to the database program allows it to apply a number of optimizations. In contrast, key-value systems treat the data as a single opaque collection which may have different fields for every record. This offers considerable flexibility and more closely follows modern concepts like object-oriented programming. Because optional values are not represented by placeholders as in most RDBs, key-value stores often use far less memory to store the same database, which can lead to large performance gains in certain workloads. Figure 1 shows an example of a Key Value Database.

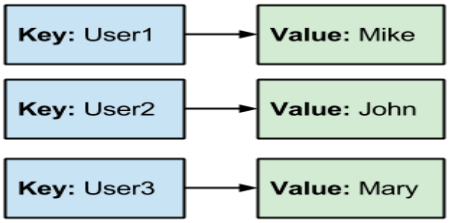


Figure 1:Key Value Database

A column-oriented DBMS is a database management system (DBMS) that stores data tables as sections of columns of data rather than as rows of data. In comparison, most relational DBMSs store data in rows. This column-oriented DBMS has advantages for data warehouses, clinical data analysis, customer relationship management (CRM) systems, and library card catalogs, and other ad hoc inquiry systems where aggregates are computed over large numbers of similar data items. An example of a column Oriented database using Customer Information is shown in Figure 2.

Customer	ZIP Code	Area Code	Sales Rep	... more columns ...	2013 Total Order
Alice	10111	212	Zack	...	$12,334
Bob	10111	763	Jane	...	$5,056
... more ...	...	...	...	...	...
Carol	52101	763	Zack	...	$2700

Figure 2:Column Oriented Database

Document-oriented databases are inherently a subclass of the key-value store, another NoSQL database concept. The difference lies in the way the data is processed; in a key-value store the data is considered to be inherently opaque to the database, whereas a document-oriented system relies on internal structure in the document order to extract metadata that the database engine uses for further optimization. Although the difference is often moot due to tools in the systems, conceptually the document-store is designed to offer a richer experience with modern programming techniques. XML databases are a specific subclass of document-oriented databases that are optimized to extract their metadata from XML documents. Graph databases are similar, but add another layer, the relationship, which allows them to link documents for rapid traversal.

Document databases contrast strongly with the traditional relational database (RDB). Relational databases are strongly typed during database creation, and store repeated data in separate tables that are defined by the programmer. In an RDB, every instance of data has the same format as every other, and changing that format is generally difficult. Document databases get their type information from the data itself, normally store all related information together, and allow every instance of data to be different from any other. This makes them more flexible in dealing with change and optional values, maps more easily into program objects, and often reduces database size. This makes them attractive for programming modern web applications, which are subject to continual change in place, and speed of deployment is an important issue.

graph database is a database that uses graph structures for semantic queries with nodes, edges and properties to represent and store data.

Most graph databases are NoSQL in nature and store their data in a key-value store or document-oriented database. In general terms, they can be considered to be key-value databases with the additional relationship concept added. Relationships allow the values in the store to be related to each other in a free form way, as opposed to traditional relational databases where the relationships are defined within the data itself. These relationships allow complex hierarchies to be quickly traversed, addressing one of the more common performance problems found in traditional key-value stores. Most graph databases also add the concept of tags or properties, which are essentially relationships lacking a pointer to another document.

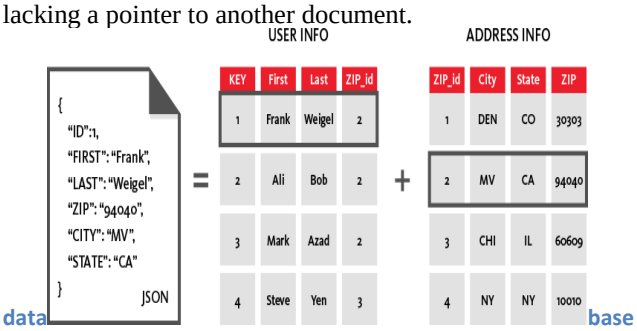


Figure 3:Document-oriented

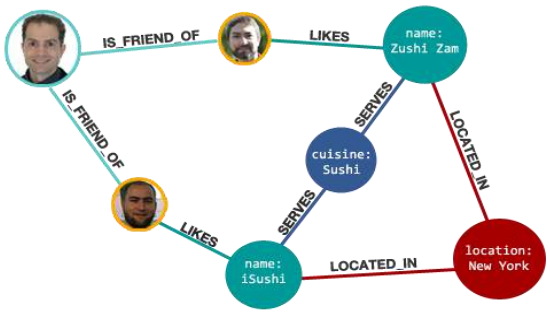


Figure 4: Graph database

III. NoSQL PROTOTYPE

A prototype system of a library inventory will be developed which will act as a platform to compare both SQL and NoSQL databases. We examine Big Data in the context of libraries. Libraries not only store huge quantities of electronic data for their product and membership information. Present-day libraries also capture Data from social media sites such as Facebook, Twitter etc. whereby library members can give feedback and ratings regarding the quality of the products and lending service.

RELATIONAL DATABASE DESIGN – SQL

A relational database organizes data in *tables* (or *relations*). A table is made up of rows and columns. A row is also called a *record* (or *tuple*). A column is also called a *field* (or *attribute*). A database table is similar to a spreadsheet. However, the relationships that can be created among the tables enable a relational database to efficiently store huge amount of data, and effectively retrieve selected data. Libraries relational database schema is designed by ER modelling. The library database has 4 many-to-many relationships, and 2 one-to-many relationships. All the tables have been fully Normalized. This process removes all redundant data from tables as it improves data integrity as well as storage efficiency. As additional table joins may be required during data retrieval therefore Normalization can impact performance.

The libraries database employs multiple table inheritance to store common attributes in a generic Product table (see Figure 5).

There are various relational databases that store data in tables and operate through queries. Oracle Application Express (APEX) was selected for implementing the libraries relational database. Oracle APEX is a freeware software development environment running inside the Oracle database.

Its web browser interface permits inexperienced programmers to quickly develop data orientated applications. APEX uses wizards and declarative programming to form powerful data entry applications, and contains a graphical query builder. Its SQL Workshop component is employed to manage database objects and run ad hoc queries.

There are several NoSQL databases that store data in JSON like documents and provide atomic scan write operations. When deciding whether or not to use NoSQL, companies

need to consider the business model, ACID transactions demand, cost and other requirements. MongoDB was chosen for implementing the libraries document information,

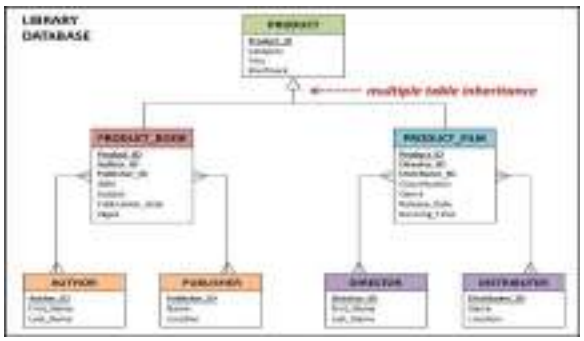


Figure 5: *Document database design - NOSQL*

Document databases store data in the form of documents that are JSON like key value pairs which are similar to rows in relational databases. Documents contain many different key value pairs, and keys may hold other key values. Documents are stored in collections; these are groups of related documents that have shared common indexes and are similar to tables in relational databases.

Document databases have a flexible schema. SQL databases requires the tables schema to be defined before inserting the data, but collections do not impose document structure. This flexibility enables the documents to match data fields of any entity, even when the data is varied substantially.

Document databases can be represented using either the reference or embedded approaches. Related Data in Embedded documents is stored in a single document. This denormalized data model allows systems to query and update related data in a single database operation.

because its flexible data model permits a prototype system to evolve during development while not modifying existing data. Its embedding model eliminates join operations and enhances query performance. it's also horizontally scalable and can reduce work by adding additional commodity servers or cloud instances. MongoDB provides a information system that stores wealthy data structures, executes complicated queries and scales out. MongoDB permits the library to insert data without a predefined schema and this is often useful for storing products with totally different attributes.

IV. IMPLEMENTATION

Firstly a relational database solution was constructed using Oracle APEX. thenceforth a document info solution was constructed using MongoDB; this style incorporates both embedding and reference data models. the aim of this section is to Illustrate the fundamental steps of each

Applications have been developed for both Android devices

implementation and highlight the significant variations between both approaches. MongoDB employs insert operations to create data; the db.collection.insert() statement adds new documents to a set. The aid field is automatically generated for a new document if the field isn't specified. The libraries document info system employs both embedded documents and references; this enables the author to demonstrate both data models.

implementation and highlight the significant variations between both approaches. MongoDB employs insert operations to create data; the db.collection.insert() statement adds new documents to a set. The aid field is automatically generated for a new document if the field isn't specified. The libraries document info system employs both embedded documents and references; this enables the author to demonstrate both data models.

```
>
> db.Products.insert({
… _id: "BCU00013",
… Category: "Film",
… Title: "Harry Potter  and the Goblet of Fire",
… Shelfmark: "COL.F02/01",
… Classification: "PG",
… Genre: "Fantasy",
… Release_Date: "November 2001",
… Running_Time:  152,
… Director:{
… First_Name:  "Chris",
… Last_Name:  "Columbus",
… Distributer:{
… Name: "Warner Brothers Pictures",
… Location: "California"}}}
```

Figure 6a: Insert document into products collections

A. EMBEDDED DATA MODEL

Embedded documents store connected data in a single document; film product documents embed the director and distributor attributes. the subsequent statement inserts a movie into the products collection (see Figure 6b):

```
>
> db.Products.insert({
… _id: "BCU0001",
… Category: "Book",
… Title: "Harry Potter  and the Goblet of Fire",
… Shelfmark: "ROW.B01/97",
… Author_id: "AUT001",
… Publisher_id: "PUB001",
… ISBN: "0-7475-6969-8",
… Subject: "Fantasy",
… Publication Date: "Nov 1994"
… Pages : 223})
>
```

V. CONCLUSION AND FUTURE WORK

Both prototype systems effectively address the libraries data storage needs, however dissent in their approach. APEX requires table structure to be defined before adding data whereas MongoDB doesn't explicitly create collections; document structure is defined automatically during the primary

data insert. MongoDB allows the library to either embed product attributes into one document or reference the data in another document. but with a normalized SQL database, product attributes may well be spread over various tables; therefore, complicated SQL joins are needed, to view all the attributes of a product. also referential integrity rules specify that the libraries SQL database should solely insert records with Author_ID into the Product_Book table, if the Author_ID exists within the Author table. The NoSQL database is very efficient and permits the library to obtain all film
attributes using one simple query. NoSQL databases make it possible to realize great

value from big data and empowers businesses to be additionally agile and scalable; this helps businesses accomplish their strategic goals and generate new revenue streams. Most new data is unstructured and the rigid schema based approach adopted by relational databases, makes it impossible to include all new data types. sensor data may be used for exploring the concept of data variety further. Future work on using NoSQL to evaluate the remaining three V's of big data is needed. Benchmark tests may be conducted to assess the handling of data volume and velocity with the deployment of NoSQL on Hadoop clusters and also the use of real-time social media data. The social media data may be in the form of text-based product reviews and also the prototype application could be extended to include sentiment analysis. NoSQL databases are becoming a worthy alternative to relational databases, as its dynamic data model is far better for managing large quantities of unstructured data; Additionally, its schema may be modified without downtime or servicedisruption.

Evaluating the Job Performance using DyScale Scheduler and MapReduce in Hadoop framework

Supriya.R, Kantharaju.H.C
Department of Computer Science and Engineering,
Vemana Institute of Technology, Bangalore-560034
rsupriya53@gmail.com, kantharajuhc@gmail.com

ABSTRACT-As there has been a massive growth in the social media in all the aspects from over a span of ten years, the amount of photos being uploaded to the Internet has been increased. Photos are being shared, downloaded and uploaded in surpass quantity through many online services like WhatsApp, Facebook, Instagram to name a few. But the applications making use of this uploaded photos are very few. Hence, in order to make use of the photos in a meaningful way, we have devised an image processing application on Hadoop framework. Image processing in this case, is used to fetch the metadata information about an image and produce the output. The framework consists of a new scheduler named DyScale using which the images are processed and compared with the fair scheduler which is usually used in Hadoop. The scheduler provides a way to use the underlying resources very efficiently and improves the performance by taking less time to process when compared to the fair scheduler.

Keywords – Hadoop, Image processing, MapReduce, Performance, Scheduling.

I. INTRODUCTION

Multi-core processing is a developing industry which replaces the single core processing system rapidly, it has the physical limits of possible complexity and speed. The present SoC (System on Chip) is designed such that it provides variety of choices in the same power envelope and help us to evaluate decision trade off. For example we can select the cores depending on the application what is being used. Hadoop is one of the most scalable and fault tolerant framework for executing bulky data sets [1]. MapReduce jobs are processed on a big cluster of service based machines. These jobs are automatically parallelized, distributed and executed over commodity machines. Hadoop was mainly designed to perform batch oriented processing of huge production jobs. The performance goals of MapReduce workloads are: large throughput oriented batch jobs and smaller response time sensitive interactive jobs. Therefore it is very difficult to take decision whether using MapReduce applications is more beneficial as compared with processors with faster cores. For example Hadoop users process a large MapReduce jobs by applying thumb rule on double size Hadoop cluster, and reduce the jobs completion in half. This thumb rule is applied on jobs, which has multiple tasks and process large data sets. The job completion time can reduce by processing these tasks on a larger number of slots. These tasks are throughput oriented and efficient processing of such jobs can be improved with additional scale out resources.The scalable operational model and data storage in Hadoop allows users to dig out information by using advanced mining technique and machine learning based algorithms to figure out novel data insights in non-traditional, game changing ways. In general, the main use of MapReduce cluster is for hosting different datasets, and multiple applications will share its compute capacity. An interesting design of heterogeneous multi-core processors [2] is to provide both fast and slow cores, for supporting different performance objectives of MapReduce jobs.

DyScale can be abbreviated as dynamic scaling scheduler with scale out or scale up approaches. Scale out approach refers to adding more servers with less RAM and processors, whereas scale up approach refers to adding more RAM and processors and also buying robust and expensive server. If MapReduce jobs are small then scale up approach can be preferred.

The paper is organized as follows. Section II briefs about the background of MapReduce, section III gives the detailed system architecture, section IV describes the image processing steps used, section V provides the results, section VI concludes the paper and gives a way to the future enhancement

II. BACKGROUND OF MAPREDUCE

MapReduce is a programming model for processing a huge amount of parallelizable data. It facilitates distributed processing of data through map and reduce stages. The map stage is partitioned into map tasks and the reduce stage is partitioned into reduce tasks. The map and reduce tasks are executed by map slots and reduce slots. Input data is processed by map function to produce data in the form of tuples and reduce function is used to combine the values along with the key. The split of input data is read by each Map task in the map stage, and applies it to generate set of intermediate key-value pairs. In the next stage map task splits and sorts these data for based on a partition function for different reduce tasks.

In the second stage, every reduce task takes its split of intermediate key-value pairs and combine these data with the same key. This method is call as shuffle or sort phase. After this, value list will be merged by using user defined reduce function to get aggregate results. This is also called as reduce phase. Output obtained from this stage will be written back on distributed file system

The job scheduling in Hadoop is generally performed as shown in the Fig 1. In Hadoop, the master node performs job scheduling, and it is called as job tracker. A Job tracker

takes request from a client and allocates with the task to the task tracker to perform. Periodically, the task tracker connects with the master job tracker and reports the present status with available slots. Based on scheduling policy and the information reported by master job tracker, job tracker decides the next job which needs to be executed. Task tracker accepts map, reduce and shuffle task from the job tracker. Job tracker keeps receiving notification from task tracker to confirm that job tracker is still active.

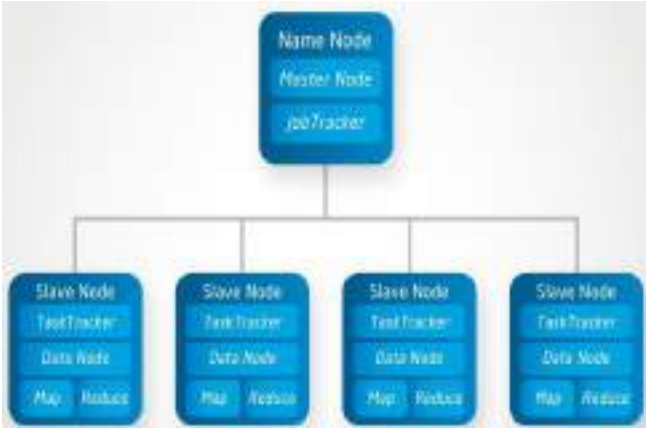


Fig 1: Job scheduling in Hadoop

III. SYSTEM ARCHITECTURE

The DyScale architecture is depicted in the Fig 2. It contains master and slave daemons. The main job of master daemon is to handle slave daemons. The slave daemon runs on every cluster nodes. DyScale executes tasks on the slaves. The master allows resources to be allocated over framework by making them available when needed. The resources like CPU, RAM contains list of Id's as (slave ID, resource1: amount1, resource2, amount2 …).

Based on organizational policy which the master chooses how many to offer in the current framework i.e. fair sharing or strict priority. To manage a diverse set of policies, master performs a plug in mechanism to utilize modular architecture. It also helps us to include new allocation modules.

There are two main components present on the framework which is running on top of DyScale which are scheduler and executer. The scheduler offers the resources and executes a framework task. It is launched on slave nodes. Master tells how many resources need to be offered to each framework and scheduler select which offered resources to be used. A description of task will be passed to DyScale when offered resources will be accepted by framework. In turn, DyScale launches the tasks on the corresponding slaves.

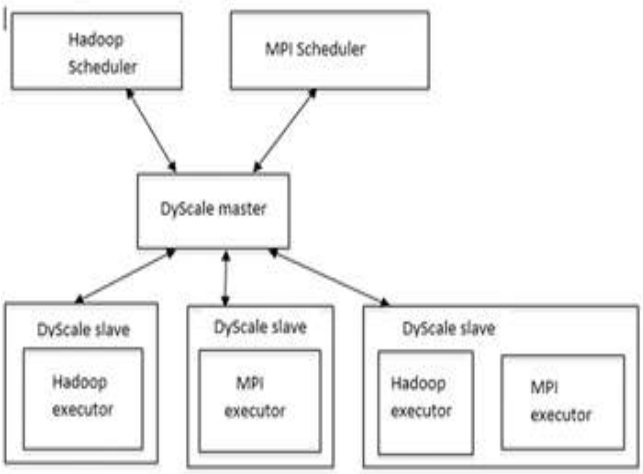


Fig 2: DyScale scheduler architecture and its components

IV. IMPLEMENTATION

In order to show the performance evaluation between the schedulers, we are incorporating image processing in Hadoop, where initially the images are downloaded from the internet, processed and the metadata is extracted from the images, then later the images are run on the schedulers in order to compare the performance. The steps in the implementation are as follows:

4.1 Hadoop Image Processing Interface (HIPI)

HIPI program considers HipiImageBundle (HIB) as a primary input object. HIB represents set of images characterized as single file on the HDFS (Hadoop Distributed File System). The distribution of HIPI contains MapReduce program that is designed on HIB from a list of images. It can be downloaded from the internet. The HIPI distribution includes MapReduce program that builds a HIB from a list of images downloaded from the Internet. The culling step is the initial stage of HIPI program. In this stage images are filtered based on different types of user defined conditions i.e. spatial resolution or criteria related to the image metadata.

The images which remain after a culling stage, are allocated to individual map task that tries to maximize data locality, which is a keystone of the Hadoop MapReduce programming model. HIPI also supports for OpenCV. OpenCV is a popular open source computer vision library. Whatever the records comes out from mapper are collected together and passed it on the reducer. Reducer uses shuffle algorithm [6] to reduce traffic on network. As the final point, all user defined reduced jobs are processed in parallel and the result is aggregated and written on HDFS. The working procedure of HIPI program is shown in Fig 3.

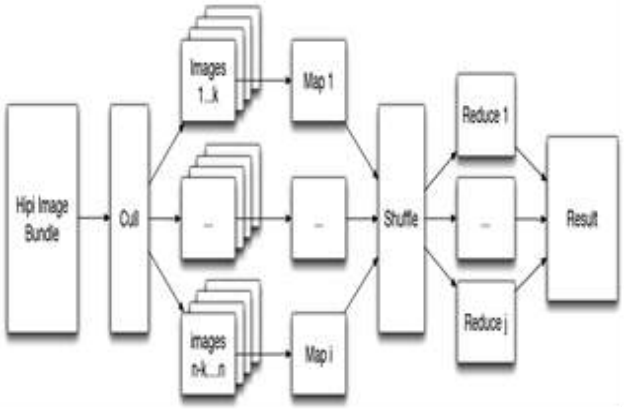


Fig 3: Organization of the HIPI program

1.2 Downloading and Storing the Image Data

In order to download and store the images, Hadoop uses the Hadoop Distributed File System (HDFS). First the image URL's list must be stored in a text file, then this image URL list is split across nodes which facilitates parallelization. Then the MapReduce will run in order to download the images from the URL list. Once it is downloaded it is stored in an image bundle known as HIB.

1.3 Processing the image bundle using MapReduce

Once the images downloaded it needs to be processed. Each individual image is processed after retrieving the image data type, after the processing the processed image is attached to the temporary image bundle where the images are stored. Each map task generates the bundles which are scattered across. In reduce phase all the scattered image files are merged into a large processed file. In order to process the images the user can use his own program based on his needs.

He needs to set the key and value as

setOutputKey (k1)

setOutputValue (v1)

Then run the below command in the terminal.

MIP-0.0.1-SNAPSHOT.jar prj.imageDataSet

which means that the user must include the specific jar file in order to process the images and the file path must be specified by the user where the images are processed and converted into the hib file and stored.

1.4 Extracting the image bundles

The framework also provides a method to extract and view the images. The extractor module is devised which extracts the images in parallel from all the available nodes. After the images are extracted identifying and arranging them in a final location is a daunting task, but our framework overcomes this difficulty by providing the user simply to extract the images in Hadoop system or in the local system. In order to extract the hib file, again the use the user should run the following command

MIP-0.0.1-SNAPSHOT.jar

MassiveImageProcessing.MIP.ImagePrs which allows the user to extract the images and store it in the path specified by the user. The MapReduce will run in order to extract the images.

2 RESULTS

After running the image dataset on two schedulers, the performance analysis graph can be plotted as bar chart shown in the Fig 4. In order to plot the graph, we should note the CPU time spent while processing the each image dataset on both the schedulers. The x-axis denotes the image dataset considered. The y-axis denotes the time taken in milliseconds. The grey colour shows the time taken by the fair scheduler to process the image whereas the blue colour shows the performance of the DyScale scheduler which takes significantly takes less time when compared to the fair scheduler which is usually used in Hadoop environment.

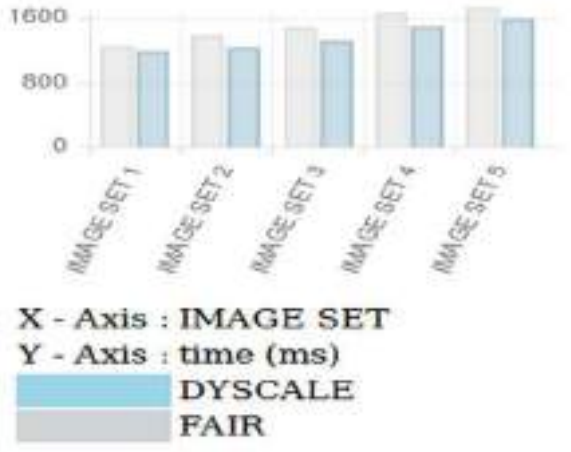


Fig 4: Performance analysis of the DyScale scheduler.

CONCLUSION

The paper can be concluded as follows, it clearly shows that the performance can be improved by using the DyScale scheduler in order to process large form of image data as it lesser time when compared to other schedulers, the dataset can be extended to other formats as well in the future. The image processing can be done considering many other use cases or criteria, also in future the process can be run on high end machines and more number of nodes in cluster while considering varied forms of input like video files and audio files.

REFERENCES

[1] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," in Proc. of OSDI '04, 2004.

[2] F. Ahmad et al., "Tarazu: Optimizing MapReduce on Heterogeneous Clusters," in Proceedings of ASPLOS, 2012.

[3] M. Zaharia et al., "Improving MapReduce performance in Heterogeneous environments," in Proceedings of OSDI, 2008.

[4] J. Xie et al., "Improving MapReduce performance through data placement in heterogeneous hadoop

clusters," in Proceedings of the IPDPS Workshops: Heterogeneity in Computing, 2010.

[5] G. Lee, B.-G. Chun, and R. H. Katz, "Heterogeneity-aware resource allocation and scheduling in the cloud," in Proceedings of the 3rd USENIX Workshop on Hot Topics in Cloud Computing, HotCloud, 2011.

[6] J. Polo et al., "Performance management of accelerated MapReduce workloads in heterogeneous clusters," in Proceedings of the 41st Intl. Conf. on Parallel Processing, 2010.

[7] W. Jiang and G. Agrawal, "Mate-cg: A map reduce-like framework for accelerating data-intensive computations on Heterogeneous clusters," in Parallel Distributed Processing Symposium (IPDPS), 2012 IEEE 26th International, May 2012, pp. 644–655.

[8] Q. Chen, D. Zhang, M. Guo, Q. Deng, and S. Guo, "Samr: A self-adaptive mapreduce scheduling algorithm in heterogeneous environment," in IEEE 10th International Conference on Computer and Information Technology (CIT), 2010.

Biographies and photographs

Ms Supriya R has received her BE degree in Computer Science from VTU, in 2014. She is currently pursuing her M.Tech degree in Computer Science in VIT, Bangalore. She has authored one journal paper and published in international conference. Her research interests include Big Data, Image processing.

Mr Kantharaju H C has received his BE degree in Information Science from VTU in 2009 and ME degree in Computer Science from Anna University in 2011. He is currently working as an Assistant Professor in VIT, Bangalore since 2013. He has authored and co-authored 5 research papers in national and international proceedings. His research interests include networking and operating systems.

Implementation of Extended MapReduce for Emerging BigData Analytics

Nayana N Kumar, Jayashree L K
Department of Computer Science,
Vemana Institute of Technology, Bengalaru-34, Karnataka
nayana.nayu14@gmail.com, jayashree_nov26@yahoo.com

ABSTRACT-Now a day's there is a need for managing huge amount of data in a faster rate is due to modern internet applications. As new upcoming data are arriving continuously, the result of data mining application becomes out of date over a period of time. This is one of the challenging jobs for computer organization to come out with new techniques and idea for handling data processing on large datasets at optimum response times. To manage vast amount of data it is required to refresh the results of mining which avoids the re computation cost from scratch. MapReduce is a technique which works based on several processors to provide automatic parallelization and distribution of computation. MapReduce framework and its open-source implementation Hadoop provides large computation environment for analysis of large-scale dataset processing and handling of dataset. This paper is focused on framework called ExtendedMapReduce, to manage the iterative operations by using map and reduce functions automatically without the user's involvement. After number of iterations small amount of updates may propagates to affect larger number of intermediate states. Which helps in improving the job running time and decreases the Incremental interactive processing time of refresh the result of big data.

Keywords –***Data mining, ExtendedMapReduce, Incremental interactive process, Hadoop, MapReduce***

I. INTRODUCTION

Many industries and organization uses big data processing technique, which is one of the trusted technologies to handle increasing amount of huge data. To improve the quality of existing services and to provide attractive new service it is important to extract important and valuable information from vast data sets, such as analysis of web-data, processing of logs and click analysis. Evaluating the huge amount of data obtained from different sources create a big challenges to the fields of science, mainly those involving massive-scale simulations and sensor networks.Modern Internet applications have created a need to manage immense amounts of data quickly. For example in social networking sites, the data produced by the user is increasing very fast every year. Big data is one of the popular techniques to take business decisions and to give better quality services. The information which is sorted and filed in the server of the organization was just data until yesterday .Suddenly the term BigData became famous and now the data filed in the company is nothing but BigData.

Involving different devices and applications data is produced which is called BigData. BigData is important in giving more accurate analysis, which may direct to more actual decision making and resulting in reduction in cost, decreased risk for business and greater operational efficiencies. To keep the extracted data up to date, periodical refreshing of mining computation is required in many situations. For analyzing BigData many frameworks have been designed. MapReduce [1] is one of the simple, generalized, framework built on Hadoop which is used for several productions. MapReduce implementation is done in large collection of computers to perform common calculations on large scale data effectively. This supports computation during hardware failure.

This paper is mainly focused on improving MapReduce technique. It supports incremental iterative process to timely accommodate new changes to the underlying data sets.

ExtendedMapReduce is advanced MapReduce technique and most sophisticated iterative computation to support key-value pair level incremental processing and extensively used in data mining applications. ExtendedMapReduce supports advanced approach called Incremental processing [2] [3], to refreshing mining results. Given the size of the input BigData it becomes very difficult to return the whole computation from scratch. Incremental processing technique takes new data from large data set, consider this as state as implicit input and combines it with new data.

II. MAP REDUCE BACKGROUND

MapReduce is one of the trusted techniques of computation; it handles large scale computations which help to stand the hardware faults. MapReduce has two main functions, called Map and Reduce. It splits the input data-set into independent chunks, and it is processed completely in parallel manner. For MapReduce computations refer Fig.1.

MapReduce supports the parallel execution, coordination of tasks that execute Map or Reduce, and also in handling one of the task that fails to execute. Key-Value pairs <K, V> sequences are turned into chunks by map task. By using input data, key-value pairs are produced, which is determined by the code written by the user for the Map function.

Each map task contains key-value pairs, are composed by a master controller and sorted by key. In reduce stage keys are divided among all, so the same reduced task wind up with same key with all key-value pairs. The reduce task processed on one key at a time and join all the values linked with that key in some way. For the reduce function the way of grouping of values is done by the code written by the user.

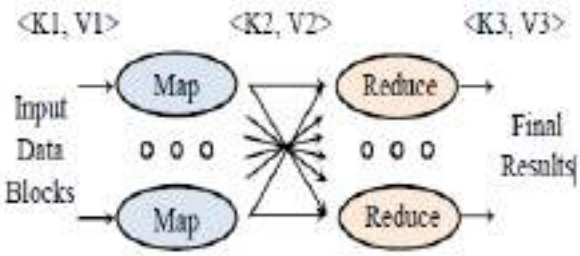


Figure 1.MapReduce computation

III. SYSTEM ARCHITECTURE

The system architecture of ExtendedMapReduce is shown in Fig.2.The implementation of ExtendedMapReduce is executed on a huge dataset and is extremely scalable. And this system can processes many terabytes of data on thousands of machines. Let us consider online data sets, by using ETL process (Extract, transform and Load), it reads the data and sends to apache Solr [4].Apache Solr is one of the open source API (Application Program Interface),is also called as NO-SQL Database, it is used as search engine for big data and accommodate data in XML(Extensible Markup Language). The performance analysis and difference with basic MapReduce and ExtendedMapReduce is shown by using K-Means Algorithm with Map Reduce Algorithm. Then for offline data set, PageRank algorithm along with MapReduce algorithm is used. For example in student database in stand Ford University the analytics is done on the student data along with iterative Algorithm implementation.

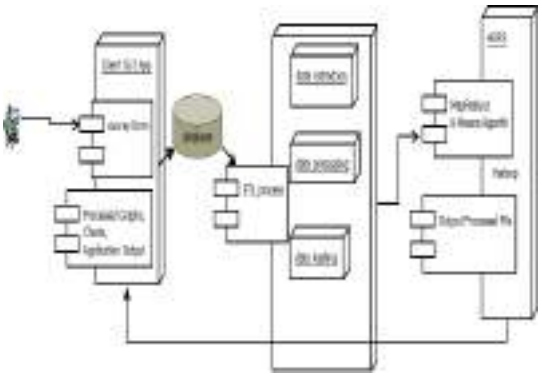


Figure 2.System Architecture

IV. IMPLEMENTATION

4.1 Analyzing Iterative Computation

PageRank [5] is a well suited web graph ranking algorithm. For each vertex in a graph ranking score will be computed by PageRank. The MapReduce job iteration is performed only after initializing all ranking scores which is shown in algorithm 1. In hyperlinked set of web pages, the PageRank calculates the numerical value for each element which reflects the probability that a random surfer will access that page.The method of PageRank can be considered as a Markov Chain [6] which needs iterative calculations to converge. Based on values calculated in the previous iteration, iteration of PageRank calculates the new access probability for each webpage. This process will continue till, the number of present iterations is larger than predefined maximum iterations, or the Euclidian distance between rank values in two subsequent iterations is less than a predefined threshold that handle the accuracy of the output results.

PageRank and MapReduce well suited with each other. Working procedure of these is as given below. It initiate with big data set called as D that has been divided into number of blocks i.e D1, D2…Dm. These blocks are circulated across different machines, such that each blocks on one machine. Let us consider Di is on machine i. These blocks are also replicated, MapReduce can ignore this. The main lamination to consider is that each machines has less memory compared to D. Now proceeds in rounds, each with 3 steps.

1. **Mapper:** Translate all d ∈ D to (key (d), value (d))

2. **Shuffle:** Moves all (k, v) and (k 0, v 0) with k = k 0 to same machine.

3. **Reducer:** Transforms {(k, v1), (k, v2) . . .} to an output D0 k = f(v1, v2, . . .).

Combiner: If one machine had multiple number of key value pairs i.e. (k, v1), (k, v2) with same key k, than reduce will perform before shuffle.

PageRank on MapReduce: v1 here is a first step. Break M into k vertical stripes M = [M1 M2 . . . Mk] so each Mj fits on a machine. Break q into q T = [q1 q2 . . . qk] (a horizontal split), again so each qj fits on a machine with M_j (This can be assumed how the data is stored, or can be done in a earlier round of MapReduce if not.) Now in each round:

• **Mapper:** j → (key= j 0 ∈ [k]; value = row r of $M_j * q_j$)

• **Reducer:** adds values for each key i to get $qi+1[j] * \beta + (1 - \beta)/n$.

The output of each mapper is considered as whole vector q_{i+1} or length n, each stripe M_j has n/k full columns. This process is feasible because qi+1 on has as many non-zero entries as Mj. However, it is not getting that much out of the combiner phase. It will see next how this can be improved.

PageRank on MapReduce: v2

Let √k and tile M into ` × ` blocks

M = M1,1 M1,2 . . . M1,` M2,1 M2,2 . . . M2,` . . . . . M`,1 M`,2 . . . M`,`

• **Mapper**: Each of k machines get one block Mi,j and get sent q_i for i ∈ [`]. • Reducer: On each row i_0 adds Mi,j qi to $q[i_0]$. Then does $q+ [i_0] = q[i_0]\beta + (1 - \beta)/n$.

4.2 K-means Clustering

K-means [7] is the most common and well used algorithm in clustering method. It takes input as parameter k and split a set of objects into k clusters. The result of intra-cluster is high and inter cluster similarity is low. The similarity of cluster is calculated using mean value of objects in the cluster.

The algorithm contains following steps. Firstly, it randomly chooses k objects from the entire objects, it represents the

initial cluster centers. Based on the distance between the object and the cluster center the remaining object is assigned to the cluster to which it is the most similar. The new mean value for each cluster is calculated. This process iterates until the criterion function converges. In this algorithm, calculation of distances is considered the most intensive calculation. For iteration it is required to compute total distance (nk), where n is a number of objects and k is the number of clusters being created. It is not relevant to compute distance between one object with the centers to the distance computations between other objects with the corresponding centers. So distance computations between different objects with centers executed parallel.

In iteration, the new centers, used in the next iteration, should be updated. Hence the iterative procedures serially executed.

To identify and handle the input and output of the implementation is the first step in designing the MapReduce for K-Means. The input is provided as a key-value pair, where 'key' is the center of cluster and 'value' is the serializable implementation of vector in the data set. Once it set the cluster and selects the centroids, and defined the data vectors that are to be clustered properly, arranged in two files then the K-Means clustering technique can be used along with Map and Reduce technique. The input directory of HDFS (Hadoop Distributed File System) prior to Map routine is contains set of centers .They form a 'key' field in the pair.

Mapper routine [8] is coded with the instruction needed to calculate the between the given data set and cluster center fed as a pair. The Mapper computes the distance between the vector value and every cluster centers mentioned in the cluster set. At the same time it also keeps track of the cluster which provide closet vector. After completion of calculation of distance, the vector is assigned to the cluster which is nearest.

It needs two main file to implement the Map and Reduce .One that houses the clusters with their centroids and another is houses the vectors to be clustered. Once Mapper is triggered the specified vector is set to the cluster that it is closest related to. After completion of this task, the recalculation is done on centroid of that, particular cluster. The reduce routine perform recalculation and prevent creations of clusters with extreme sizes (cluster having too less data vectors or a cluster having too many data vectors) by restructuring the cluster At the final stage, the centroid of the given cluster is updated, and re written the new set of vectors and clusters on disk which is ready for the next iteration. After understanding of what the input, output and functionality of the Map and Reduce routines it design the Map and Reduce classes by following the algorithm discussed below.

Algorithm 1. map (key, value)

Input: Global variable centers, the offset key, the sample value

Output: pair, where the key' is the index of the closest center point and value' is a string comprise of sample information

1. Construct the sample instance from value;
2. minDis = Double.MAX VALUE;
3. index = -1;
4. For i=0 to centers.length do dis= ComputeDist(instance, centers[i]); If dis < minDis { minDis = dis; index = i; }
5. End For
6. Take index as key';
7. Construct value' as a string comprise of the values of different dimensions;
8. output < key , value > pair; 9. End

Algorithm 2. reduce (key, V)

Input: key is the index of the cluster, V is the list of the partial sums from different host Output: < key , value > pair, where the key' is the index of the cluster, value' is a string representing the new center

1. Initialize one array record the sum of value of each dimensions of the samples contained in the same cluster, e.g. the samples in the list V ;
2. Initialize a counter NUM as 0 to record the sum of sample number in the same cluster;
3. while(V.hasNext()){ Construct the sample instance from V.next(); Add the values of different dimensions of instance to the array NUM += num;
4. }
5. Divide the entries of the array by NUM to get the new center's coordinates;
6. Take key as key';
7. Construct value' as a string comprise of the center's coordinates;
8. output < key , value > pair;
9. End

V. RESULTS

Finally comparing difference between Normal MapReduce and ExtendedMapReduce for different applications like K-means clustering and PageRank the Overall performance analysis of ExtendedMapReduce is as shown in Fig. 3.In every iteration, map function avoids reading and parsing the structure data by splitting the structure and state data. With incremental processing, the performance will improved in ExtendedMapReduce, so reducing the Normal MapReduce by 98%. When compared to Normal MapReduce, the advanced ExtendedMapReduce shuffles and input changes will affect the intermediate kv-pairs from the Map instances. Thereby further improving the shuffle time, achieving 95% reduction of NormalMapReduce time. For the sort stage, ExtendedMapReduce sorts the little number of kv-pairs from the changed Map instances, thus eliminating almost all sorting cost of NormalMapReduce. For the Reduce stage, iterMapReduce cuts the run time of Normal MapReduce by 80% because it does not require combining the updated state data and the structure data.

Figure 3 Performance of ExtendedMapReduce.

CONCLUSION

The ExtendedMapReduce is advanced MapReduce technique that supports the iterative processing for large datasets and give solution to the issues which arises during MapReduce implementation of iterative processing. It offers an environment and a model to the programmers for designing and to perform explicitly for iterative algorithms and moreover suggesting the concept of persistent tasks to carry out the iterative computation to keep away from continually creating, destroying, and scheduling tasks. In same iteration, it can also perform asynchronous execution of tasks, to accelerate the processing speed.

FUTURE ENHANCEMENT

As a result of the incremental processing, the MRBGraph file may contain multiple segments of sorted chunks, each resulting from a merge operation. So this situation needs to be improved and there is a need to enhance the query algorithm with a multi-window technique to efficiently process the multiple segments.

REFERENCES

[1] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *in Proc. of OSDI '04, 2004.*

[2] D. Peng and F. Dabek, "Large-scale incremental processing using distributed transactions and notifications," in *Proc. Of OSDI '10, 2010, pp. 1–15.*

[3]P.Bhatotia, AWider R and R.Pasquin,"Incoop:Mapreduce for incremental computations",in *Proc. Of SOCC '11,2011*

[4] Apache giraph. http://giraph.apache.org/.

[5] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. Mc-Cauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributeddatasets: A fault-tolerant abstraction for. in-memorycluster computing," in *Proc. of NSDI '12, 2012.*

[6]S. Ewen, K. Tzoumas, M. Kaufmann, and V. Markl, "Spinning fast iterative data flows," PVLDB, *vol. 5, no. 11, pp. 1268–1279, 2012.*

[7] Y. Zhang, Q. Gao, L. Gao, and C. Wang, "imapreduce: A Distributed computing framework for iterative computation," J. Grid Comput., *vol. 10, no. 1, 2012.*

[8] Y. Bu, B. Howe, M. Balazinska, and M. D. Ernst, "Haloop: efficient iterative data processing on large clusters," PVLDB, *vol. 3, no. 1-2, pp. 285–296, 2010.*

Biographies and Photographs

Nayana N Kumar received the B.E. degree in 2014 from Visvesvaraya Technological University. Currently, she is an M.Tech. Candidate in computer science, Vemana IT Bengaluru Karnataka. Her research interests include Data mining and BigData processing.

Jayashree L K is an Asst. professor at Vemana IT. And working from past 15 years in Dept. of CSE, Vemana IT Bengaluru Karnataka. Her current research interests are data ware housing and analysis of BigData.

A Friend Recommendation System based on Similarity Metric and Social Graphs

Rashmi. J, Dr. Asha. T
Department of Computer Science
Bangalore Institute of Technology, Bangalore, Karnataka, India
rash003.j@gmail.com, asha.masthi@gmail.com

ABSTRACT-All the existing social networking services recommend friends to users based only on their social graphs, which is not very appropriate in reflecting user's preferences in selecting a friend in real life. In this paper, we present a friend recommending system for social networks, which recommends friends to users based on their life styles as well as social graphs, as the proposed friend recommending mechanism is being integrated into social network. By taking the advantage of sensor-rich smart phones, the proposed system discovers the life styles of users , measures the similarity of life style existing between the users, calculates the friend recommendation score using the proposed similarity metric, and recommends friends to query user who are having high friend recommendation scores. Since the proposed system is integrated into the social network, the existing feature of social network i.e. the social graphs is used for recommending social friends to the user. Therefore on receiving the request, the proposed system returns a list of people with high friend recommendation scores as well as a list of social friends to the query user.

Keywords – *Friend Recommendation, Life Style, Social Graphs, Social Networks*.

I. INTRODUCTION

Twenty years ago, people typically made friends only with the people who lived close to themselves such as neighbors or colleagues. The friends made through this fashion are termed as G-Friends, which stands for geographical location-based friends as they are influenced by the geographical distances between them. The rapid advances in the social networks, services such as Facebook, and Twitter have provided us revolutionary ways for making friends. According to the statistics of the Facebook, a user has an average of 130 friends [1].

One challenge residing in the existing social networking services is recommending a good friend to the user. Most of the existing friend recommending systems relies on pre-existing user relationships to suggest friend candidates. For example, Facebook relies on social graphs to recommend friends to the user, i.e. users who share same geographical location or same profession are recommended as friends to the user, which is not very appropriate in reflecting user's preferences in selecting a friend.

According to the studies [2] and [3], the basic rules for grouping people together are: 1) life styles;2) attitude;3) interests;4) moral standards;5) economic level;6) already known people. Most of the existing friend recommendation systems consider rule #3 and #6 as the main factors for recommending friends to users. Our proposed system considers rule #1, #3 and #6 as the main factors for recommending friends to users. Life styles are correlated with daily routines and activities performed by the people. The life style of the people comprises of activities such as shopping, travelling, playing sports, swimming, listening to music, watching TV etc. This proposed friend recommendation mechanism is deployed as an add-on to the existing social networking services, hence making it as a hybrid friend recommendation system which utilizes both the social graph feature of the existing social networking service and the similarity metric feature of the proposed system.

II. LITERATURE SURVEY

Recommendation systems that suggest items to the users have become popular in the recent years. For example, Amazon [4], recommends items to the user based on their previous visit and the items that are frequently visited by the other users. Netflix [5] and Rotten Tomatoes [6] recommend movies to the users' based on previous users' ratings and habits of watching.

Over the recent years, with the advances in the social networking services, friend recommendation has gained a lot of attention. The existing friend recommendation systems like Facebook and Twitter recommend friends to user based on their social relations.

In the meantime many other recommendation systems have been proposed by researchers. Bian and Holtzman [7] have presented a collaborative friend recommendation system called as MatchMaker that is based on personality matching. Kwon and Kim [8] have presented a friend recommendation system that is based on physical and social context. But the authors have not explained what a physical social context is and how to obtain that information.

These existing friend recommending systems are different from our proposed system. In our work, we exploit the recent sociology findings to recommend friends based on their similar life styles as well as social relations. The advance of smart phones enables activity recognition using the set of sensors on smart phones.

III. SYSTEM OVERVIEW

This section gives the high-level overview of the friend recommendation system. Fig. 1 shows the architecture of the proposed friend recommendation.

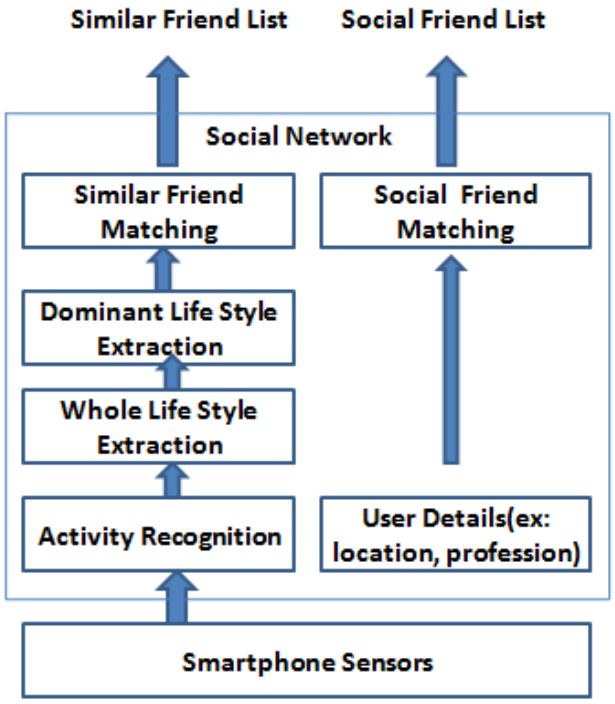


Fig. 1. System architecture of Friend Recommendation system.

In the activity inference phase, the activity of each user is recognized that is collected from the smartphones. The activities of the users are collected for a certain period of time. In the life style extraction phase, the users' whole life style and the dominant life style are extracted. From the activities recognized in the activity inference phase, the whole life style of the users are extracted i.e. the set of activities that are performed both frequently and infrequently in a given period of time, and are added to the MySQL database.

The whole life style activities are then given as input to the apriori algorithm which then computes the frequently performed activities that represents the dominant lifestyle for the given user and . In the friend matching phase, the dominant life of the query user is compared with all the other users and the no. of matching activities are compared, and using the proposed similarity metric a friend recommendation score is computed.

The computed friend recommendation score for each user exceeding the defined threshold value represents a friend to the query user with high similar life style. In the social friend matching phase, the profession and the geographical location details of the query user is compared with other users, the users details matching with the query user are recommended as social friends to the query user as they are social related. The following sections will elaborate on all the modules of the proposed system.

3.1 ACTIVITY RECOGNITION

The life styles are a mixture of motion activities performed by the user in the daily life. The sensors on the smart phone are used for inferring user's motion activity. Since the number of activities involved in the analysis is unpredictable, unsupervised learning approach is used for organizing the activities. K-means clustering algorithm is used for grouping data into clusters, each cluster representing an activity. Since the raw data collected by the smart phones are noisy, median filter is used for filtering the noisy data. The cluster centroids are calculated and distributed to the smartphones. The smartphones then recognize the activity based on the minimum distance rule and uploads the sequence of activity to the server instead of raw data.

We have the implemented the activity recognition phase of the proposed as a website consisting of several urls'. Here the urls' represent the activities performed by the user in the daily life. Here we have considered activities like shopping, travelling, listening to music, watching TV, cooking etc. Each url is represented using a integer. The url and its associated integer value is added to MySQL database. The users registered with application can login to this website. Once the user logs into the website, he/she visits the url of his/her choice. An activity of the user is recognized when he/she visits the url, representing an activity or set of activities performed by the user in his/her daily life.

The following table shows how the url and its associated integer value is stored in database.

Table. 1. Activities and their corresponding id's stored in the database.

Activity_id	Activity
1	http://www.soundcloud.com
2	http://www.imdb.com
3	http://www.google.com/shopping
4	http://www.booking.com
5	http://www.webmd.com
6	http://www.howstuffworks.com
7	http://www.wired.com
8	http://www.yahoo.com/tech
9	http://www.indeed.co.in
10	http://www.bigfishgames.com
11	http://www.allrecipes.com
12	http://www.break.com

3.2 WHOLE LIFE STYLE EXTRACTION

Since life style is a combination of activities performed by the user in his/her daily, in our implementation urls' visited by the user in the given session represents the life style of the user. In real life, the activities of the user are observed for certain number of days. In our implementation, the activities are tracked for many sessions, so that the life style of the user can be predicted accurately. The urls' representing the activities of the user, when visited by the

user is added to the database along with its session id. This is done for all the users for tracking their life style. The life styles tracked in the above specified way are termed as whole life style of the user, as they are a combination of both frequently and infrequently performed activities.
The following table shows how the whole life style of each user is stored in the database.

Table. 2. Whole life style activities of each user stored in the database.

User_id	Tracked_activities	Session_id
22	1-6-3-4-5	00233A78EE888888B88113
31	1-9-2	0456CC899AABB22899EE81
28	2-4-5	06445EE733666FF556A77BC
24	2-6-7-8	100125672891AADDD7777B
30	1-2-4-5	167755AABBEE8902263772
22	2-3-4	186777DDBBA8999E77655F
23	1-3-4	198767E5678F7666A77889B
25	2-5-8-9	20008953738EFD67A89B998
22	1-4-5-6	2001037467624882784FFED
26	2-4-5-6	21763758632BAECFFDE8488
25	1-3-5-7-8	222873485678EFAD8745876

3.3 DOMINANT LIFE STYLE EXTRACTION

To calculate the similarity of life styles between the users, only the whole life style activities of the user cannot be used, as they are a combination of both frequently and infrequently performed activities. To determine the dominant life style of the user, only the activities performed frequently by the user must be considered. Hence the dominant life style of each must be computed. Once the whole life style of the user is obtained, those set of activities are given as input to the Apriori algorithm. The application of the Apriori algorithm is to compute the frequent set of items i.e. the set of items occurring frequently for the given set of items. In the proposed system, the whole life style is treated as the given set of items, which then computes the frequently occurring item sets i.e. in the proposed system the algorithm computes the activities that are frequently performed in a given period of time. The set of frequently performed activities obtained represent the dominant life style of the user. We have considered a support of 30% in algorithm for computing the frequent item sets i.e. the frequently performed activities.

i. Apriori Algorithm

- For each item,
  - Check if it is a frequent itemset
  //appears in > minimum support transactions
  - k=1
  - repeat
    //iterative level-wise identification of frequent itemsets.
  - For each new frequent itemset Ik with k items
    //level k+1
  - Generate all itemsets Ik+1 with k+1 items, Ik is a subset Ik+1
  - Scan all the transactions once and check if the generated k+1 itemsets are frequent
  - k=k+1
  - Until no frequent itemsets are identified.

The following screenshots show how the dominant life style for one user is calculated.

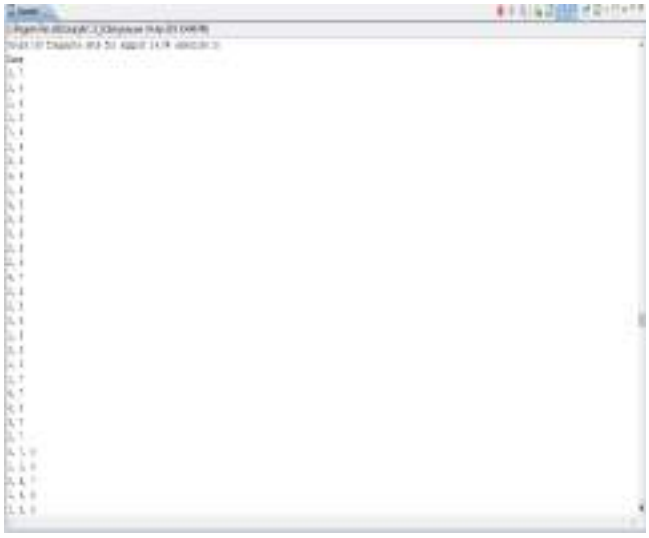


Fig. 2. Frequent set of activities being computed using Apriori algorithm.

Fig. 3. Dominant life style computed for the given user using Apriori algorithm.

The following table shows the computed dominant life styles for all the users.

Table. 3. Dominant life style computed for each user stored in the database.

User_id	Dominant_life_style
22	1 ,2 ,5 ,6, 7
23	1 ,2 ,5
24	1 ,2 ,3 ,5 ,6
25	1 ,2 ,3 ,5
26	1 ,2 ,5 ,6, 7
27	2 ,6 ,7
28	1, 5, 6, 9
29	3, 5, 6, 7
30	2, 6, 7, 8
31	1, 3, 4, 5
32	8 ,9, 11

3.4 SIMILAR FRIEND MATCHING

Once the dominant life style of the all the users are obtained by the Apriori algorithm. The dominant life styles of all users are compared with query user's dominant life style. From the life style comparison, parameters like the no of activities matching with each user and total life style match value are obtained.

The proposed similarity metric computes the friend recommendation score for each user using the above values obtained on comparison. A threshold value is defined for the friend recommending system. The list of users whose friend recommendation scores exceed the predefined threshold value are recommended as friends sharing similar life style with the query user. Here we have defined the threshold value as 4. Hence all the users' friend recommendation scores exceeding 4 are recommended as friends sharing similar life styles. The friends' list contains only the names of the users, to preserve the privacy of the users' by not revealing the users' life style details.

The friend recommendation score is computed using the following the equation:

F_score=matching activities + whole life style match (1)

Where

- F_score: friend recommendation score
- Matching activities: no of activities between the query user and the user considered for friendship.
- Whole life style match: this value is 1 if all the activities match in the life style set matches otherwise zero.

The following table shows the friend recommendation scores that are computed for all the users.

Table. 4. Friend recommendation score computed for each user stored in the database.

User_id	No_of_activities_matching	Whole_life_style_match	Friend_recommendation_score
23	3	0	3
24	4	0	4
25	3	0	3
26	5	1	6
27	3	0	3
28	3	0	3
29	3	0	3
30	3	0	3
31	1	0	1
32	0	0	0

The following screen shots show how friends with similar life styles are recommended.

Fig. 4. Screenshot showing the friends' list sharing similar life style with the query user.

3.5 SOCIAL FRIEND MATCHING

Social graphs represent the social relationship existing between the people in the graph.The people who share social relations are termed as social friends. Social relations are based on the profession, geographical location, etc. that the people share with others. Already known people are also termed as social friends. Recommending friends to users based on the social relationships is the feature of the existing social networks. Facebook and Twitter also relies on the social graphs for suggesting friends to the users. Since we are incorporating the proposed friend recommending mechanism into the social networks, we are making use the existing social graphs feature for suggesting the social friends to the users along with the friends sharing similar life style. Hence the proposed system behaves as a hybrid friend recommendation system recommending both similar life style friends as well as social friends to the query user.

The following screenshots depict how the proposed system recommends social friends to the query user.

Fig. 5. Screenshot showing the profile details of the query user.

Fig. 6. Screenshot showing the social friends' list for the query user .

CONCLUSION

In this paper, we have presented the design and implementation of a friend recommendation system that is based on similarity metric and social graphs. The proposed system behaves as a hybrid friend recommendation system, recommending both social friends and friends' sharing similar life style to the query user, as it is incorporated in the social networking service. Hence the user is provided with a wide range of choices for selecting a friend for his/her preference. Also privacy is preserved, which is achieved by revealing only the names of friends in the friend list and not their life style details to the query user.

In future, the activities of the users representing their behavior can be kept tracked at the server/admin side. Therefore, if any user is involved in any activities such as crime, then it can be easily identified by their activities that are observed and stored at the server/admin side.

REFERENCES

[1] Facebook statistics. (2011). [Online]. Available: http: // www.digitalbuzzblog.com /facebook-statistics-stats-facts-2011/

[2] G. Spaargaren and B. Van Vliet,"Lifestyle consumption and the environment : The ecological modernization of domestic consumption," Environ. Politic, vol. 9, no. 1, pp. 50–76, 2000.

[3] M. Tomlinson, "Lifestyle and social class," Eur. Sociol. Rev.,vol. 19, no. 1, pp. 97–111, 2003.

[4] Amazon. (2014). [Online]. Available: http:// www.amazon.com/

[5] Netflix. (2014). [Online]. Available: https: //signup-netflix.com/

[6] Rotten tomatoes. (2014). [Online]. Available: http: //www.rottentomatoes.com/

[7] L. Bian and H. Holtzman, "Online friend recommendation through personality matching and collaborative filtering," in Proc.5th Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol.,2011, pp. 230–235.

[8] J. Kwon and S. Kim, "Friend recommendation method using physical and social context," Int. J. Comput. Sci. Netw. Security,vol. 10, no. 11, pp. 116–120, 2010

Authors Profile

Rashmi.J is currently pursuing M.Tech degree from Bangalore Institute of Technology, Bangalore. She has obtained her B.E degree from B.T.L. Institute of Technology and Management, Bangalore. Her research interests are Social Networking and Web Services.

Dr. Asha.T obtained her Bachelors and Masters in Engineering, from Bangalore University, Karnataka, India. She has her Ph.D from Visveswaraya Technological University under the guidance of Dr. S. Natarajan and Dr. K.N.B. Murthy. She has over 20 years of teaching experience and currently working as Professor in the Dept. of Computer Science & Engg., B.I.T. Karnataka, India. Her research interests are Data Mining, Medical Applications, Pattern Recognition and Artificial Intelligence.

A Selective Approach for Storing Small Files in Respective Blocks of Hadoop

Chethan.R , ChandanKumar, Jayanth Kumar.S Girish.H.J , Prof. Mangala.C.N
UG Scholar, Associate Professor
Department of Computer Science and Engineering
chethanr95@gmail.com, kchandan488@gmail.com,sjayanth317@gmail.com,girishgowdahj@gmail.com
mangalacn.02@gmail.com

Abstract **–** Hadoop is an open source framework used for processing the data in big data. This hadoop majorly consists of two components 1. HDFS(Hadoop Distributed File System) 2. Map Reduce Component. HDFS is used for storing the files in hadoop and Map Reduce is used to process the data stored in HDFS. Hadoop doesn't performs well for storing the small files that is, it provides individual block of DataNode to individual file and hence reduces the performance. This research work gives an introduction of HDFS and the existing ways for solving the problem of small files. In this proposed approach, we merge the small files of into same block using Map Reduce programming model on Hadoop and hence provide different key value for files of different format. Hence this approach reduces the inefficient usage of memory from NameNode to access the DataNode and in turn it improves the efficiency of Hadoop by storing selective small files in respective blocks of Hadoop.We also propose a Traffic analyser with MapReduce paradigm that provides batch analysis in minimum response time and helps to process the log files in efficient and stable way

Keywords - Hadoop; NameNode; MapReduce; Small Files; Traffic Analyzer.

I. INTRODUCTION

It is a known fact that Hadoop has been specially created to manage Big Data. We know that Google is the world known popular search engine. To provide search results for users, Google had to store huge amount of data. Hence in 1990's,they started searching for the different ways of storing these huge amount of data and finally in the year 2003 they came up with Google File System(GFS) to store these huge amount of data and in 2004 they provided another technique called MapReduce for processing the data present in GFS.

But these techniques were presented to the world as a description and was just stored theoritically in GFS. So people had knowledge of the technique but there was no working model or code provided. Then in the year 2006 another major search engine, Yahoo came up with techniques called HDFS and MapReduce based on the descriptions given by Google. So, finally HDFS and MapReduce became the two core components of Hadoop.

Hadoop was actually created by Doug Cutting. Doug Cutting[2][7] choosed the logo of hadoop as an elephant. The reason behind it is that, the elephant is symbolic representation and a good solution for Big Data.

This paper covers many sections. Section II covers about the Hadoop distributed file structure, and the MapReduce component. Section III discusses the small file problems and the existing approach. Section IV discusses the proposed approach. Section V covers experimental setup. Section VI discusses conclusion and future work and then the paper is concerned with the acknowledgement for the constant support provided to us.

II. HADOOP COMPONENTS

The two main components of Hadoop are HDFS and MapReduce.HDFS(Hadoop Distributed File Structure) is used for storing the files in Hadoop and also consists of two nodes called NameNode and DataNode to split the given data into blocks and then store them in respective blocks. MapReduce component is used to process the data stored in the HDFS that is it involves in processing large amount structured and unstructured data in parallel in order to maintain good performance for the system. The three major components of this Hadoop architecture is as shown below

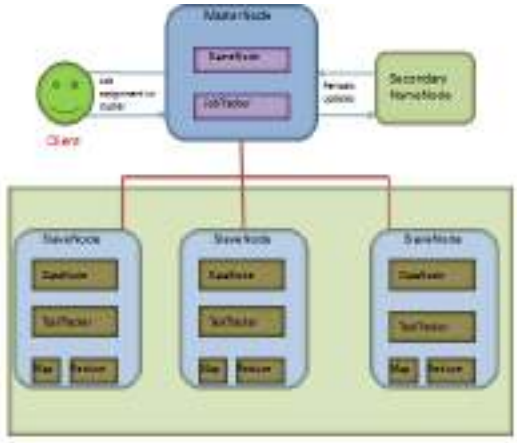


Figure 5: Hadoop Distributed File System Architecture

A.NameNode

NameNode is an center piece of an HDFS file system. It keeps the directory tree of all the system and tracks where across the cluster file data is kept and it does not store the data of these files itself.Client applications talk to NameNode whenever they wish to locate a file,or when they want to add/copy/move/delete a file.The NameNode responds the successful requests by returning a list of relevant DataNode servers where the data lives.Hence these nodes are called Master Nodes.These also consists JobTracker which is a deamon runs on the Name Node.The secondary NameNode is connected to the NameNode which access acts a backup by storing replicas of the metadata of the file systems in local storage.

B.DataNode

A DataNode are refered as slave nodes since they follow the commands given by the NameNodes. They store data in the

Hadoop File System.A functional filesystem has more than one

DataNode, with data replicated across them. On startup,a DataNode connects to the NameNode;spinning until that service comes up and then responds to requests from the NameNode for filesystem operations. Client applications can directly talk to the DataNode,once the NameNode has provided the location of the data. TaskTracker is a deamon process on the DataNode which indeed should be deployed on the same servers that host DataNode instances.

C.HDFS Client

HDFS Clients are neither master nor slave,

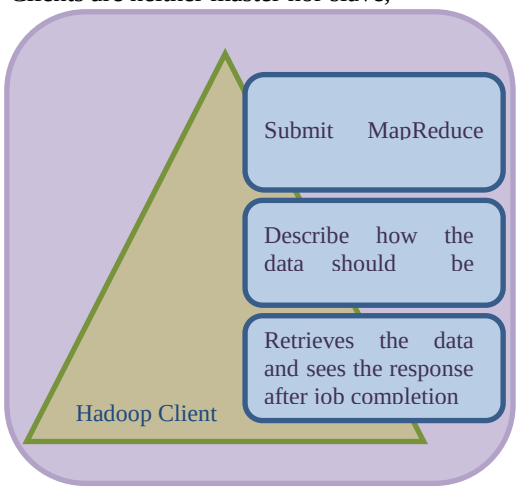


Figure 2:HDFS Client structure

rather play role of loading the data into cluster, submit MapReduce jobs describing how the data should be processed and then retrieve the data to see the response after the completion of the given job.

III. SMALL FILE PROBLEMS IN HADOOP AND EXISTING APPROACHES

The Hadoop Distributed File system is a distributed file system mainly designed for batch processing of large volume of data. The default block size of HDFS is 64MB. Storing lot of small files which are extreamly smaller then the block size cannot be efficiently handled by HDFS. When data is represented in files significantly smaller than the block size the performance degrades dramatically[5]. Mainly there are two reasons for producing small files.One reason is some files are pieces of a larger logical file.Other reason is some small files cannot be combined together into one larger file and are essentially small.When small files are used there will be lots of seeks and lots of hopping from DataNode to DataNode to retrieve each small file which is an inefficient data access pattern[6]. Hadoop offers few options to handle these small files problems.They are as follows:

1) CONSOLIDATOR

Consolidator takes a set of files containing records belonging to the same logical file and merges the files together into larger files. It is possible to merge all small files into one large file,but it is not practical as then it would be a terabytes sized file.It would take a longer time to run such a huge file.

2) HAR FILES[6][9]

Hadoop Archieves were introduced to HDFS to alleviate the problem of lots of files putting pressure on the NameNodes memory. HAR files work by building a layered file system on top of HDFS.

A HAR file is created using the hadoop archieve command,which runs a MapReduce job to pack the files being archieved into a small number of HDFS files. The below figure 3 shows the architecture of HAR which is containing two index called *Index* and *MasterIndex*..Reading through files in HAR is comparatively slower than reading the files in HDFS because HAR files requires two 'index' file reads as well as the data file read.

This is one of the disadvantage of HAR file.

In order to overcome the disadvantage of this HAR the locality of speed in HAR should be improved in order to increase the speed of access.

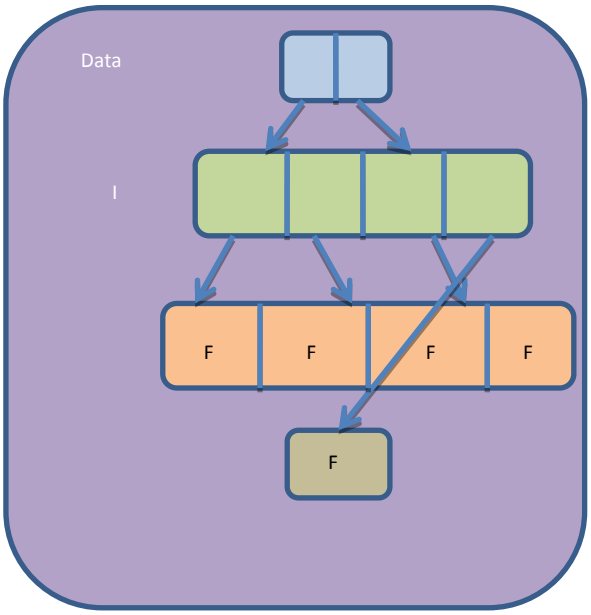


Figure 3: HAR File Layout

3) USING HBASE STORAGE

HBase stores data in MapFiles(indexed Sequence Files) and therefore it is a good choice when it is need to do MapReduce streaming analysiswith the occasional random look up. But the major problem is it doesn't allows partial keys completely and allows only one default sort per table.

4) SEQUENCE FILES[6][9]

Sequence files in a Hadoop specific archieve file format to tar and zip. The below figure shows Sequence file layout and the concept behind this is to merge the file set with using a key and a value pair and this created files are known as 'Hadoop Sequence Files'. In this method file name is used as key and file content is used as value but it is very much time consuming to convert existing data into sequence files.

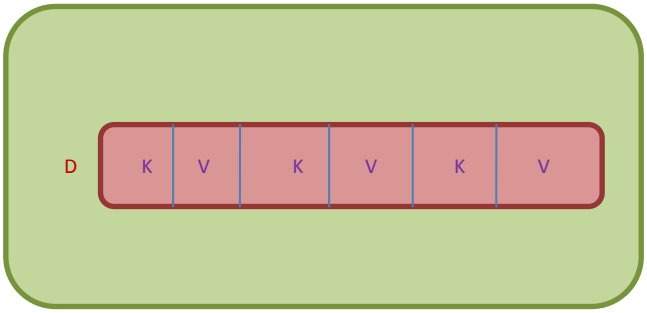


Figure 4: SequenceFile File Layout

IV. PROPOSED APPROACH

In this section, details of the proposed approach are introduced. Initially the idea of the approach is defined and the the algorithm and the mathematical model is described.

Storing large number of small files into HDFS is an overhead in terms of memory usage of NameNode and increase in executing time of MapReduce. According to this problem analysis, the proposed approach was merging the selective small files into respective blocks of Hadoop and hence make them a large file. Hence this will reduce the number of files and saves the memory of HDFS. Before merging all the small files together we determine the files of same format stored in the blocks of Hadoop, then these small files can be combined in parallel using MapReduce paradigm where Mapper will fetch the file and during these mapping technique the Mapper should provide a key and value for these small files inorder to fetch them. In this approach Mapper is made to provide keys for the file which is the byte offset and the value for these files will be the filename. Now the Mapper starts adding the files until it reaches the default block size and then pass it to reducer. The reducer will merge the files. This process is then carried out parallel until all the files are completely merged. This approach will reduce the time required for merging and executing the files and also makes it easier to access a file of particular format in the whole blocks of Hadoop. This approach also reduces the time of execution by ignoring to merge those files whose size is more than the threshold which is set to 80%(0.8) of the block size of Hadoop. This threshold can also be given as an input to the algorithm and it should be a integer number between range 0 to 1.

Algorithm for Map in MapReduce

i. Identify and fetch the small files and place them in a block of Hadoop knowing their filename and filesize.

ii. Consider the respective filename as key and filecontent as value for all the individual files in the block.

iii. If the file size is greater than the threshold ignore to add in the list.

iv. Maintain a list of block name(key given to block) and file names for merging which is to be done by reducer.

v. Pass this entire list to Reducer.

vi. After this the list becomes empty and fetches the new input file.

Algorithm for Reducer in MapReduce

i. Take the input from Mapper.

ii. Merge the files considering the threshold.

Mathematical Model

Let K_r be the set of Keys in $U^{`r}$, let $V_r$ be the multiset of values in $U^{`r}$, and let $(V_{k,r})$ denote the multiset of values in $U^{`r}$ that have key K. Then K_r and V_r $\theta(n^{1-\varepsilon})$ can be partitioned across machines such that all machines get $O(n^{1-\varepsilon})$ bits, and the pair $(k, V_{k,r})$ gets sent to the same machine.

Proof : For a set of binary strings *B* denoted by *s(B)=bЄB |b|* the total space used by the strings in *B*. Since the algorithm is in MRC, by definition,

$s(V_r) + s(K_r) \leq s(U^{`r}) = O(n^{2-2\varepsilon})$

Furthermore , the space of the reducer is restricted to $O(n^{1-\varepsilon})$; therefore ¥ K, |K|+s($V_{k,r}$) is $O(n^{1-\varepsilon})$.

We can conclude that maximum number of bits mapped to any one machine is no more than the average load per machine plus the maximum size of any $(k, V_{k,r})$ pair.

Thus,

$$\leq \frac{s(V_r)+s(K_r)}{\text{number of machines}} + \max_{K \in K_r}(|K|+s(V_{k,r}))$$

$$\leq \frac{O(n^{2-2\varepsilon})}{\theta(n^{1-\varepsilon})} + O(n^{1-\varepsilon})$$

$$\leq O(n^{1-\varepsilon})$$

V. EXPERIMENTAL ANALYSIS

In this paper, we use the experiment of WordCount to test the performance HDFS using MapReduce in processing and storing the small files. This experiment can be performed for data of any format as of now we are providing the setup only for text file and also we compare the processing time required by proposed approach and the previous existing approaches for the same experiment.

WordCount

WordCount is a problem which is used to determine the count of repeatation of the words in the given file or data. The experiment was conducted by creating a text document and we appended it with some texts like "*hello how are you hope you are good*" and with some more text but the above text shows the data that was split and stored into a block of HDFS.

MapReduce consists of 3 components Mapper, Shuffel and Combine, Reducer. Now this data is sent to Mapper by NameNode and and Mapper takes two input *Key* and *Value* and returns two outputs that is the same *Key* and *Value*.Now Mapper divides each word and text into a single split and each splits are executed parallely. The Key for Mapper is the byte offset and the value is the content.

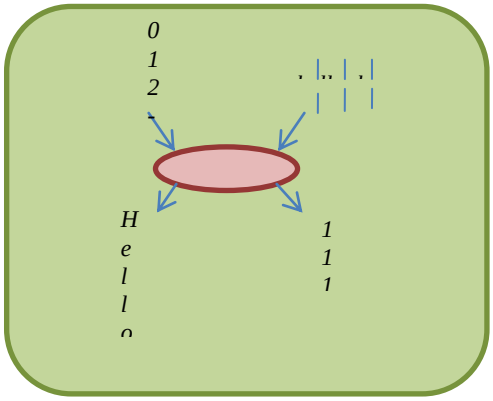


Figure 5: Mapper function

The value 1 which is given as an output indicates the index value for the corresponding splits.Now this output is given to the Shuffle and Combiner of MapReduce which combines the index value of the repeated words into a single block

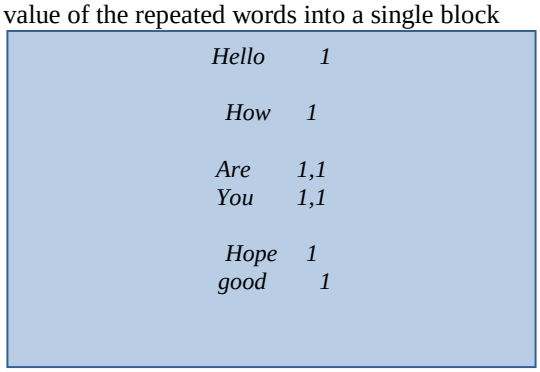


Figure 6: Shuffel and Combine function

After the completion of the shuffle and combine task the obtained output is sent to the last component of MapReduce that is Reducer which counts the number of index values and replace it by corresponding value and hence this provides the total number of count of a particular word in a given text file for the *WordCount* problem. The below figure shows the function performed by the reducer after taking the input from the Shuffel and Combine component.

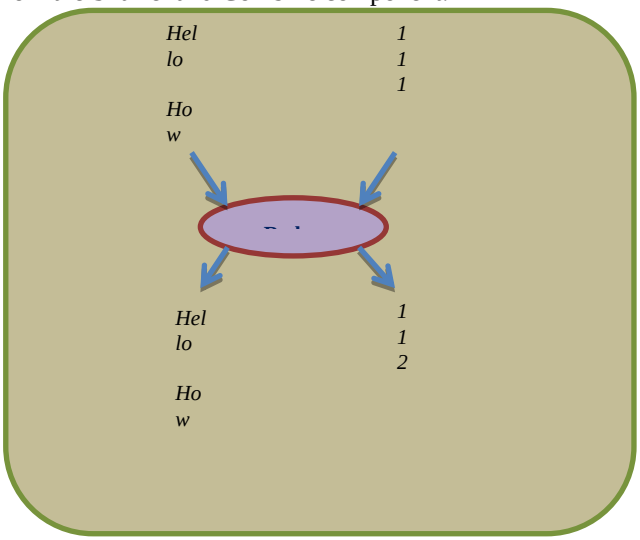


Figure 7: Reducer function

Once after performining the experiment with the proposed approach and the existing approaches we compared the execution time of all these approaches using a graph and then we came to the conclusion that the proposed system can perform more well in terms of execution time and the memory management.

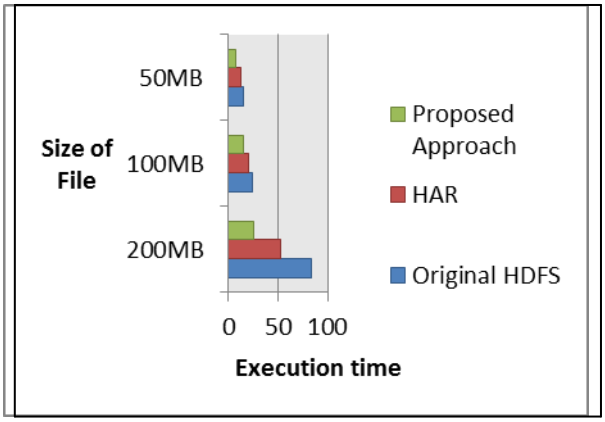


Figure 8: Comparision of different approaches of execution time on different sized files

The above graph in figure 8 depicts that the proposed approach takes less amount of time to execute the small file. In HAR approach there is inconsistency with the block size. Sequence file takes lot much of time to text data to sequence file format therefore we have not considered this approach and Original HDFS also takes lot of time to execute the application for small files and the proposed approach takes less amount of time for its execution since all the splits given to the Mapper by the NameNode are executed parallel and hencr reduces the number of Mappers which in turn increases the performance of the Hadoop by decreasing the execution time required for processing the small files.

VI. CONCLUSION AND FUTUREWORK

Hadoop is being one of the wide area of research in handling of small files in HDFS, hence the following research focuses on MapReduce approach to handle the small files and retrieve them using the key values given to the merged files. The proposed approach also focuses on execution time to run small files on Hadoop Cluster and hence the performance of HDFS. This can handle both sequence file and files of text, pdf etc which is related only to text file efficiently and also avoid files whose size is greater than threshold.

In future, work can be carried out to other files like audio,vedio and image files which are also a kind of small file. These files can be stored in HDFS and they also suffer performance issues which were faced with the small files discussed previously.

ACKNOWLEDGMENT

We would like to thank Mrs.Mangala.C.N, Associate Professor, Department of Computer Science and Engineering, East West institute of technology for her constant support and her guidance and also for providing us the opportunity to contribute on the small file problems faced by Hadoop. At last we also thank all faculty members of Department of Computer Science and Engineering for helping us in this work.

REFERENCES

[1] Manghui Tu, Peng Li, I-Ling Yen, BhavaniThuraisingham, Latifur Khan.Secure Data Objects Replication in Data Grid., IEEE Transactions on Dependable and Secure computing, Vol. 7, No. 1,2010

[2] Apache Hadoop, http: //hadoop.apache.org/

[3] P. H. Cams, W. B. Ligon III, R. B. Ross, and R. Thakur, PVFS: A parallel file system for Linux clusters, in Proc. of 4th Annual Linux Showcaseand Conference, 2000, pp. 317327.

[4] Hong Sha, Yang Shenyuan, Key technology of cloud computing and research on cloud computing model based on Hadoop, Software Guide,2010,9(9): 9 I I.

[5] Xie Guilan, Luo shengxian, Research on applications based on Hadoop MapReduce model, Microcomputer & its applications, 2010, (8): 4 7.

[6] The-small-files-problem, Cloudera http://blog.cloudera.comlblog/ 2009/02/the-small-files-probleml, February 2009

[7] Dhruba Borthakur, The Hadoop Distributed File System: Architecture and Design, ACM 2006

[8] Amazon Web Services, http://aws.amazon.comlec 2/, 2006

[9] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler "The Hadoop Distributed File System" IEEE 2010

[10] Techniques-to-deal-with-small-files-in.html, http://hadoopshares.blogspot.in /2012/09/techniques-to-deal-with-smallfiles- in.html, September 2012

AUTHORS BIOGRAPHY

Mrs.Mangala.C.N received the B.E degree in Computer science and Engineering from NCET, Bangalore, VTU University in 2006 and got M.Tech degree in Computer Science from RVCE, Bangalore, India. She is currently working as Associate Professor in the faculty of CSE,EWIT-Bangalore, India. Her area of interest includes Image Processing, Data Mining and Big Data.

Mr.Chandan Kumar is persuing his B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Big Data, Data Mining and Network Security.

Mr.Jayanth Kumar.S is persuing his B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Big Data, Data Mining and Cloud Computing.

Mr.Girish.H.J is persuing his B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Big Data, Cloud Computing and Network Security.

Mr.Chandan Kumar is persuing his B.E in Computer Science and Engineering in East West institute of Technology, Bangalore, India. His area of interest includes Big Data, Data Mining and Cloud Computing.

Integration of Visual Temporal and Textual Distribution Information for News Video Mining

Prof Shivamurthy R C, Tauseef Ahmed S S
Department of Computer Science, Akshaya Institute of Technology, Tumkur
mail4tauseef@gmail.com, shivamurthyrc@gmail.com,

ABSTRACT- News web videos exhibit several characteristics, including a limited number of features, noisy text information, and error in near-duplicate key frames (NDK) detection. In this paper, a novel framework is proposed to better group the associated web videos to events. First, the data preprocessing stage performs feature selection and tag relevance learning. Next, multiple correspondence analysis is applied to explore the correlations between terms and events with the assistance of visual information. Finally, a probabilistic model is proposed for news web video event mining, where both visual temporal information and textual distribution information are integrated. . Co-occurrence and visual near duplicate feature trajectory induced from NDKs are combined to calculate the similarity between NDKs and events. It is needed urgently the advanced technologies for organizing, analyzing, representing, indexing, filtering, retrieving and mining the vast amount of videos to retrieve specific information based on video content effectively, and to provide better ways for entertainment and multimedia applications.

Keywords - Co-occurrence, multiple correspondence analysis (MCA), near-duplicate key frames (NDK), news web video event mining, trajectory.

I. INTRODUCTION

Advances in the media and entertainment industries, including streaming audio and digital TV, present new challenges for managing and accessing large audio-visual collections. Search engines and video sharing websites such as YouTube, YouKu, Google, make it convenient for the users to access relevant news web videos. News wires like CNN, BBC, and CCTV also publish news videos. 2014data show that over 100 h of videos are uploaded to YouTube every minute, and six billion hours of videos are watched each month on YouTube. These facts demonstrate a new challenge for the users to grasp the major events available from searching video databases. To address this need, news web video event mining approaches have been developed. When the users search a topic, most want to know: what happened, why it happened, and how it happened. Major event mining can facilitate more effective news web video browsing and a better understanding of the entire topic through the relationships among events. For example, sample search results of the topic "London terror attack" from YouTube are demonstrated in Fig. 1. The results are mainly ranked by text relevance or popularity, which means that a thumbnail image and its corresponding sparse set of tags are not sufficient to help users understand the main content of the topic. The users have to browse many news web videos in the returned list and even watch most of the videos. This is not only time consuming, but also difficult if thousands of news web videos are returned by a search engine. This situation calls for effective approaches to automatically group relevant news web videos into events, and then mine the relationships among them Visual information suffers from semantic gap and user subjectivity problems, and textual information can be noisy, ambiguous, and sometimes incomplete. Therefore, using either visual or textual information alone for news web video event mining usually yields unsatisfactory results. Such challenges motivate us to utilize both visual and textual features for news web video event mining, with the aim to overcome their shortcomings, while leveraging their advantages.

For visual information, machine learning of visual and text features to perform supervised annotation and some important shots are frequently inserted into videos or reports as a reminder or support of viewpoints, which carry useful information. Because of the unique role of near-duplicate key frames (NDK) in the news search, topic detection and tracking (TDT) and copyright infringement detection, these duplicate shots/key frames are clustered to form different groups according to visual content.

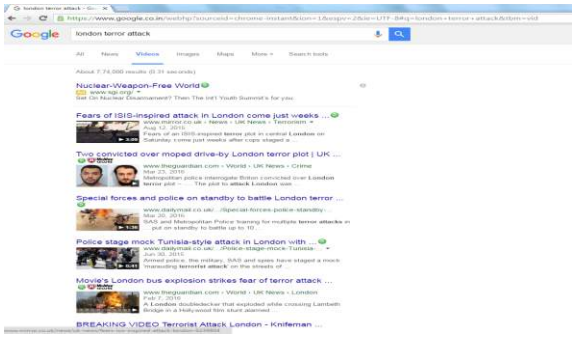


Fig. 1. Search results of "London terror attack" in YouTube on April, 2016.

Visual near-duplicate feature trajectory is an extension of the textual feature trajectory, which models the visual feature distribution along the timeline in a 2-D space with one dimension as time, and the other as the feature weight. In this paper, the visual features are referred to the NDKs. It

can evaluate the importance of an NDK varying over time, which will be used to cluster those NDKs belonging to the same event with different viewpoints. In another way, NDK-within-video information can enhance the robustness of the visual near-duplicate feature trajectory, while it is affected by the NDK editing/detection problem. Therefore, we explore integrating NDK-within-video information and visual near-duplicate feature trajectory, as visual temporal information, to cluster more NDKs belonging to the same event. In contrast with broadcast news videos, where there is textual information from speech transcripts for event mining, news web videos generally have much less textual information (such as titles and tags). This is because news web videos are usually uploaded by users who do not make rich titles and tags for the videos. In addition, these titles and tags are typically noisy, ambiguous, and incomplete. Even worse, the users may include irrelevant hot terms (words) simply to attract attention. Because of the different backgrounds or habits, people use a wide variety of terms to describe the same video. This characteristic would impact the textual distribution information (i.e., the term weight of each word in each NDK) in multiple correspondence analysis (MCA). Therefore, we explore using NDK neighbors to enhance the weights of the terms, which would better represent the high level semantic information of NDKs. In this paper, a novel framework is proposed that integrates the visual temporal information and textual distribution information for news web video event mining. After feature selection and tag relevance learning by neighbor voting, MCA is explored to extract the NDK-level event similarity with the assistance of textual information. Next, both co-occurrence information and visual near-duplicate feature trajectory induced from NDKs are used to detect the similarity between NDKs and events. Finally, in order to integrate visual and textual information for event mining, a hybrid probabilistic model is proposed. Although some of the techniques adopted in this paper, such as NDK, MCA, and visual feature trajectory have been used in previous work, here we bring these techniques together to complement each other. The main novelty and contributions of this paper are as follows.

1) To address the uncontrolled user tagging, ambiguity, and Over personalization, both the MCA similarity measure and neighbor stabilization process (visual neighbor information) are integrated to generate the textual distribution information, which accurately and efficiently learns tag relevance by the visual-content relationship between NDKs to improve the robustness of the textual information. Moreover, it can better explore the degree of correlation between different terms and events.

2) The visual near-duplicate feature trajectory, i.e., the time distribution information of an NDK, is integrated with the NDK-within-video information (co-occurrence) as the visual temporal information to cluster more NDKs belonging to the same event, which obtains robust and accurate features to improve web event mining.

3) A novel unified probabilistic framework is proposed to integrate the visual temporal and textual distribution information for news web video event mining.\

II. RELATED WORK

2.1. Definitions

Topic is a seminal event or activity, along with all directly related events and activities. Therefore, we can infer that atopic consists of events and activities. A "hot topic" in is defined as a topic that appears frequently over a period of time. Generally, a "hot topic" has the following characteristics:

(1) It appears in many news stories on one or more news channels;

(2) It has a strong continuity, which means that many different events relevant to the topic are also reported; and

(3) Its popularity changes over time. An event in topic detection and tracking (TDT) is defined as something that happens at a specific time and place, along withal the necessary preconditions and unavoidable consequences. Such an event might be a donation, a game, or a concert performance. An activity in TDT is the connected series of events with a common focus or purpose, which happens in specific places during a given time period. For instance, inactivity may be a campaign, a survey, or an earthquake relief.NDK is a group of key frames that are visually similar, but appear different because of the variations introduced during the acquisition time, lens setting, lighting condition, and editing operation. NDKs have been used in the real-world applications such as TDT.

2.2. Topic Detection and Tracking

TDT automatically structures online news articles into topics. TDT detects new topics and tracks known events in text news streams, and hence, many studies inset focused on text data. Studies have been conducted in the multimedia field. The topics were tracked with visual duplicates, which resulted in the concept of NDK. The novelty and redundancy detection was explored in, in which visual duplicates and speech transcripts are integrated to measure the similarity of cross-lingual news stories. With the assistance of NDK constraints, news stories were clustered into topics by constraint-based co clustering. News stories from different TV channels were linked by textual correlation and key frame matching. The retrieval of NDKs plays an important role in measuring the video-clip similarity and tracking video shots of multilingual sources. The system in, segmented news videos into stories and constructed the dependencies among stories as a graph structure. The interface Media Walker supported in browsing the

development of news topics. With the textual correlation and key frame matching, topic clusters were grouped in and news stories from different TV channels were linked in .A video log management model was proposed in, which is comprised of automatic log annotation and user-oriented log search. Sports video semantic event detection was explored in, which is based on the analysis and alignment of the webcast text and broadcast video. Experiments demonstrated that the incorporation of webcast text into sports video analysis significantly facilitates the sports video semantic event detection. Topic discovery was deployed by constructing the duality between stories and textual visual concepts through bipartite graphs. Visual-text time-dependent alignment was explored in to summarize the topics. Text co-occurrence and visual feature trajectory we refused for news web video event mining.

2.3. Feature Trajectory

Feature trajectory is a statistical measure for the information retrieval. It evaluates the importance of a feature which varies over time. The characteristics of word trajectory were analyzed in to identify the important and less-important, periodic and periodic words, from the perspective of time-series word signal. A parameter-free probabilistic model was proposed to analyze the time-varying features and to detect the burst events from text streams. The idea of mining hot terms by the timeline analysis was presented in. Hot topics were further extracted using multidimensional sentence modeling grounded on hot terms. GoogleTrends was used to predict the milestone events of a topic.

2.4. Multiple Correspondence Analysis (MCA)

Multiple Correspondence Analysis (MCA) can measure the correlations among multiple variables, which is able to capture the correlation among nominal variables. The feature-value pairs and classes can be projected into a 2-D space constructed by the first and second principal components, because of the fact that over 95% of the total variance can be captured by the first two principal coordinates. The function of MCA is to use the textual distribution information to mine the correlation among NDKs and events. However, the characteristics of the textual information (including noisy, ambiguous, incomplete, synonyms, polysemy, and Multilanguage) of news web videos make news web video event mining a challenge problem. The property of the textual information would affect the efficiency of MCA. Therefore, we explore using NDK neighbor information to get more robust and effective semantic relationships among NDKs. MCA and co-occurrence were applied to NDK-level news web video event mining through linear integration. We have extended the study in several ways. First, in the current work, the visual near-duplicate feature trajectory is integrated with the NDK-within-video information (co-occurrence) as the visual temporal information to cluster more NDKs belonging to the same event. Second, for the textual distribution information, we explore using the associations among NDKs to find more related terms (term group) of each NDK. Then we try to make use of the semantic relationships between term groups and events to mine the similarity between an NDK and an event. Third, as the input of MCA, the indicator matrix between text terms and NDKs is in the form of binary values. In this current work, a weighted matrix is adopted to model the distribution of each term in NDKs more accurately.

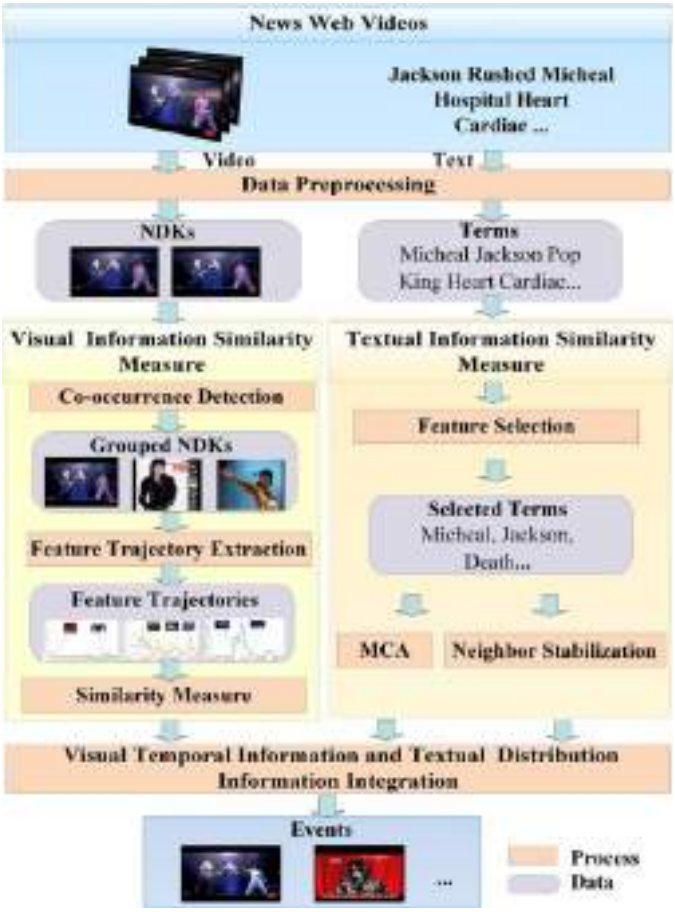


Fig. 2. Proposed framework for news web video event mining.

III. NEWS WEB VIDEO EVENT MINING

This consists of four stages: data preprocessing, textual information similarity measure, visual information similarity measure, and hybrid probabilistic model for integration (see Fig. 2). After getting the correlation between terms and events through visual neighbor information, MCA is used to calculate the similarity between each NDK and event through textual distribution information in NDKs. For the visual information, NDK-within video information and visual near-duplicate feature trajectory are fused to improve the robustness and accuracy of the visual features. Finally, the content-based visual temporal information and textual

distribution information are integrated through the proposed probabilistic model. In this study, event mining is achieved within each topic. The input of our framework is the news web videos returned from a user query. After NDK detection and grouping, a series of NDKs and their corresponding terms can be obtained. As a result, the similarity between each NDK and each event is calculated, and every NDK is assigned to the event with the largest similarity. The output is the classified events.

3.1. Data Preprocessing

After shot boundary detection, the middle frame in each videos hot is extracted as the key frame for the shot. Each video can be represented by a series of key frames. Then, NDK detection method is utilized to detect the NDKs among videos. Local points are detected with Harris–Laplace and described by SIFT. The detected NDKs are further grouped to form clusters by transitive closure. For each topic, (1) is applied to calculate the probability of an NDK belonging to each event

$$P(\text{NDK}p,Eq) = |\text{NDK}p \_Eq| / |\text{NDK}p| \quad (1)$$

where $|\text{NDK}p \_Eq|$ is the number of common videos between NDKp and event Eq . $|\text{NDK}p|$ is the number of videos whose key frames are in NDKp. Eq is the qth event. Finally, each NDK is marked as the event label which has the largest probability. Ground truth is manually determined according to the search results from Wikipedia and Google. Terms extracted from titles and tags are treated as textual features. Because of the noisy user-supplied tag information, text words are pruned using methods, including word stemming and special character removal.

3.2. Textual Information Similarity Measure

Feature selection and visual neighbor information are used to enhance the weights of the representative terms. The less important terms are neglected. MCA is then applied to calculate the MCA-based transaction weights, targeted to bridge the gap between an NDK and terms. In order to apply MCA, each feature in the training data set is discredited into several partitions (i.e., feature-value pairs), and the same partition ranges are used to discredited the testing data set. As a result, the similarity between each feature-value pair and an event is calculated. Finally, the weights between each NDK and all events are calculated by summing the weights of the feature-value pairs along all features.

3.3. Visual Information Similarity Measure

NDK-within-video information is first used to measure the similarities among NDKs. Second, the visual feature trajectories induced from NDKs are used to find the highly relevant NDKs as the time distribution feature. Because of the complementary characteristics of the NDK-within-video information and the time distribution information, both are utilized to measure the similarity between an NDK and an event.

3.4. Hybrid Probabilistic Model for Integration

A hybrid probabilistic model is proposed for better video event mining, which integrates the visual and textual information. Ultimately, every NDK is grouped to the event with the largest similarity value.

IV. TEXTUAL INFORMATION SIMILARITY MEASURES

Generally, news web videos use titles and tags to describe their content. The features extracted from terms are treated as textual information. There are numerous frequently accompanied terms from titles and tags, which convey useful information. Hence, we propose to mine the textual distribution information as a part of our proposed framework.

Fig. 3. Each video is described by its title/tag, which is taken as textual information. The features of each term extracted from this textual information are taken as the tag-based features.

4.1. Feature Selection and Neighbor Stabilization

For feature selection, the chi-squared statistics with respect tithe classes are adopted to evaluate the importance of the terms with WEKA. Noisy terms would impact the accuracy of MCA. Therefore, word pruning is an unavoidable problem. All terms are ranked in descending order, and then the gaps between neighbors in this sorted list are calculated. Significant terms are ranked above the largest gap based on our empirical study. Finally, the largest gap is used as the

threshold, where all the terms with smaller weights are filtered.

4.2. Multiple Correspondence Analysis with Correlation Information

MCA is composed of two steps: training and testing. In order to illustrate the principle of MCA, the training process is taken as an example. First, all the features are combined to Forman indicator matrix with NDKs as the instances (rows), terms and event labels as the categories of variables (columns) (settable I), where in the testing process, it does not need to label the NDKs. After calculating the tiff value of each term distributed in each NDK (textual distribution information), this indicator matrix can be represented in a 2-D table NT (NDKs versus Terms)

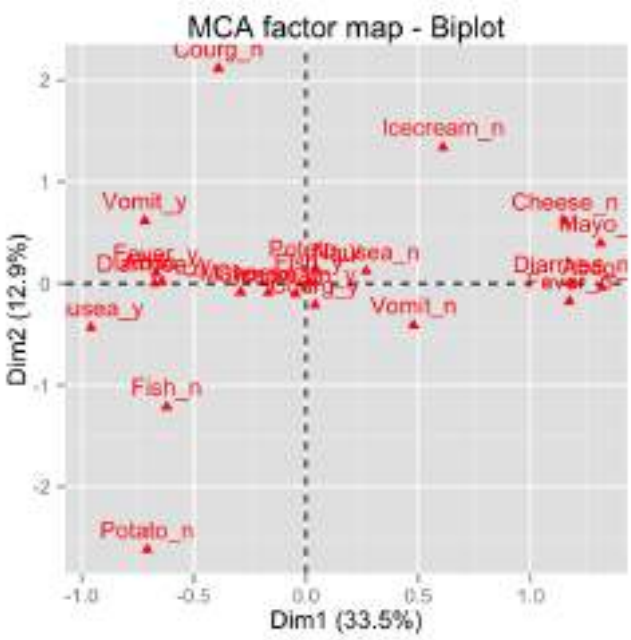


Fig. 4. Geometrical representation of MCA.

4.3. Textual Similarity Measures

We calculate the similarities between each NDK and all events using

$$\text{Sims}(NT_k, E_r) = \gamma \times TW_{k,r} + (1 - \gamma) \times \text{Sim}_T(NDK_k, E_r).$$

V. VISUAL INFORMATION SIMILARITY

It is indeed easier to know whether two videos are similar or not rather than to obtain their true labels, because the space of potential labels is very large.

News web videos have fewer textual features than text documents, and these features are often noisy, ambiguous, and incomplete. Thus, video content information compensates for the textual information. The NDKs are taken as the visual "terms, "which contain rich information. Since NDKs provide a strong cue to link event-relevant videos across sources, languages, and times, an NDK supports key frame redundancy. The features derived from NDKs are treated as visual information, such as NDK-within-video information and visual near duplicate feature trajectory.

5.1. Burstyn Period Detection

A web search engine usually returns a large number of search results, mainly according to the text relevance. Some results may not be relevant. For example, when searching "Michael Jackson Dead," most users want to view news web videos from the accident on June 25, 2009. However, because of the keyword "Michael Jackson," a large number of news web videos about "Songs of Michael Jackson" will also be listed. Actually, each topic appears frequently within a certain period. To discover the events from the search results, it is essential to locate the burst period according to the video upload time. The method is used to locate the burst region of each topic as follows:

$$Raj = [tj - w, tj + w]|Vj| \geq \alpha\_n\ k=1(|Vk|)\ n$$

5.2. NDK inside Video Information

Important visual shots are frequently inserted into the relevant videos. These NDKs usually carry useful video content information and can be used to group videos of similar themes into events. There are numerous frequently accompanied NDKs that convey useful information. As shown in Fig. 3, each NDK is composed of a series of key frames, such as NDK1=_*V*1 1, *V*2 1 _ and NDK2=_*V*2 2, *V*3 2 _.

Both NDK1 and NDK2 are appeared in video V2.

Data preprocessing is the first stage. Multiple Correspondence Analysis (MCA) is then applied to explore the correlation between terms and classes, targeting for bridging the gap between NDKs and high-level semantic concepts. Next, co-occurrence information is used to detect the similarity between NDKs and classes using the NDK-within-video information.

5.3. Visual Near-duplicate Feature Trajectory

The visual near-duplicate feature trajectory models the visual feature distribution along the timeline in a 2-D space (one dimension as time and the other as feature weights). The importance increases proportionally to the number of videos appearing in a certain time-period. Usually, a few representative NDKs will often appear in a period of time, but rarely in other time periods. These NDKs with similar trends imply their consistency. Hence, the relevance of these visual feature trajectories can be clustered to form events.

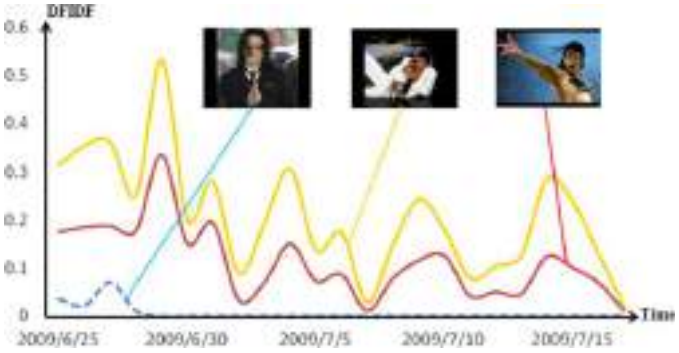


Fig. 5. Different NDKs on the same event could have similar visual near duplicate feature trajectories. For example, the key frames of "Jackson is dancing in an MTV" and "Cover of MTV" share similar visual near-duplicate feature trajectory distributions over time. They belong to the event "A tribute of Michael Jackson dead." In contrast, the key frame of "Jackson is praying" shows different feature distributions, which belongs to the event "Sadness of Michael Jackson dead."

5.4. Visual Similarity Matches

Different NDKs on the same event could have similar visual feature trajectories. For example, both the events "A tribute of Michael Jackson dead" and "Sadness of Michael Jackson dead" belong to the same topic (i.e., "Michael Jackson dead").As shown in Fig. 5, the NDK of "Jackson is dancing in an MTV" and "Cover of MTV" share similar visual near-plicate feature trajectory distributions over time, since they belong tithe same event "A tribute of Michael Jackson dead." Incontrast, the key frame of "Jackson is praying" shows a different feature distribution, while it belongs to another event "Sadness of Michael Jackson dead." Similar scenes might have multiple trajectories because of the NDK detection error, video editing, or other reasons. Some NDKs belonging to one group may be falsely detected to form several separated clusters. Moreover, anew NDKs are missed and falsely treated as non-NDK. Here the current visual feature trajectories deviate from the ideal one. Fig. 6shows an example of several NDKs on "Jackson is dancing in the last rehearsal." Unfortunately, they to the same event "Last Rehearsal" with similar scenes, but demonstrate different trends. Since co-occurrence can further group NDKs with relevant visual content, we believe that it can enhance the robust and consistent characteristics of the visual near duplicate feature

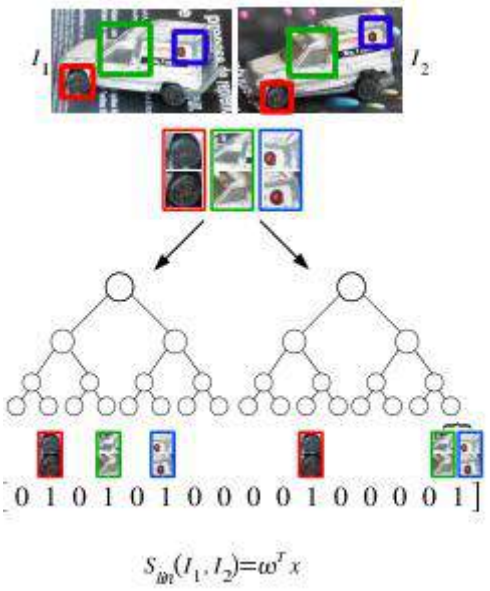


Fig. 6. Example of similar scenes but with multiple trajectories.

trajectories and, thus, improve the performance of news web video event mining.

VI. CONCLUSION

In view of the unique characteristics of news web videos, such as the limited number of features, the unavoidable error in NDK detection, and noisy text information, news web video event mining has been a challenging task. In this paper, a novel hybrid probabilistic framework is proposed for news web video event mining, which integrates the textual and visual information, and aims to solve not only noisy and limited textual information but also the unavoidable video editing and NDK detection problems. Next, a visual neighbor information extraction method proposed to deal with the well-known ambiguity and overly personalized problems. Meanwhile, the purgative textual information helps to bridge the gap between NDKs and the high-level semantic concepts. Moreover, both the textual and visual features with relatively low frequencies are considered a useful information in our experiments.

REFERENCES

[1]Statistics. (2014). [Online]. Available: http://www.youtube.com/yt/Press/statistics.html

[2] J. Yuan, Y.-L. Zhao, H. Luan, M. Wang, and T.-S. Chua, "Memory recall based video search: Finding videos you have seen before based on your memory," *ACM Trans. Multimedia Comput., Commun. Appl.*,
vol. 10, no. 2, pp. 1–21, 2014.

[3] X. Wu, C.-W. Ngo, and Q. Li, "Threading and autodocumenting newsvideos: A promising solution to rapidly browse news topics," *IEEE SignalProcess. Mag.*, vol. 23, no. 2, pp. 59–68, Mar. 2006.

[4] Y. Ke, R. Sukthankar, and L. Huston, "Efficient near-duplicate detectionand sub-image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2004,pp. 869–876.

[5] C. Zhang, X. Wu, M.-L. Shyu, and Q. Peng, "A novel web video event mining framework with the integration of correlation and co-occurrenceinformation," *J. Comput. Sci. Technol.*, vol. 28, no. 5, pp. 788–796, 2013.

[6] K.-Y. Chen, L. Luesukprasert, S.-C. T. Chou, "Hot topic extraction based on timeline analysis and multidimensional sentence modeling," *IEEETrans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1016–1025, Aug. 2007.

[7] C.-W. Ngo, W.-L. Zhao, and Y.-G. Jiang, "Fast tracking of near-duplicatekey frames in broadcast domain with transitivity propagation," in *Proc.14th ACM Int. Conf. Multimedia*, 2006, pp. 845–854.

[8] X. Wu, C.-W. Ngo, and A. G. Hauptmann, "Multimodal news story clusteringwith pairwise visual near-duplicate constraint," *IEEE Trans. Multimedia*,vol. 10, no. 2, pp. 188–199, Feb. 2008.

[9] J. Cao, C.-W. Ngo, Y.-D. Zhang, and J.-T. Li, "Tracking web video topics:Discovery, visualization, and monitoring," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 21, no. 12, pp. 1835–1846, Dec. 2011.

A SURVEY OF VARIOUS TECHNOLOGIES FOR SOA ADOPTED BY SOFTWARE INDUSTRIES

Karthik .V, Prof. R.Jaya
P.G. Scholar, Sr. Assistant Professor, Department of Computer Science and Engineering,
New Horizon College of Engineering, Bangalore, Karnataka, India
karthik.v114@gmail.com, jayamanojkumar@gmail.com

ABSTRACT-This paper presents the knowledge about technologies in Service- Oriented Architecture (SOA) and Web services. To integrate the various applications of a large enterprise, we need various standards to remove the heterogeneity problems. The various standards and its responsibilities are depicted in this paper. The standards for discussion are SOAP, WSDL, UDDI, XML, and JAVA. Finally, we depict the technologies and development challenges and directions in the field of SOA and Web services.

Keywords - Service integration, SOAP, UDDI, WSDL, XML.

1. INTRODUCTION TO SOA AND WEB SERVICES

In an SOA, applications are made up of loosely coupled software services, which interact to provide all the functionality needed by the application. Each service is generally designed to be self-contained and stateless to simplify the communication that takes place between them. There are three major roles involved in an SOA:[1]

1.1: Service provider
The service provider creates a service and can publish its interface and access information to a service broker.
A service provider must decide which services to expose and how to expose them. Often, a trade-off exists between security and interoperability; the service provider must make technology decisions based on this trade-off. If the service provider uses a service broker, decisions must be made about how to categorize the service, and the service must be registered with the service broker using agreed-upon protocols.[1].

1.2: Service broker
The service broker, also known as the service registry, is responsible for making the service interface and implementation access information that is available to any potential service requester.

The service broker provides mechanisms for registering and finding services. A particular broker might be public (for example, available on the Internet) or private, only available to a limited audience (for example, on an intranet). The type and format of the information stored by a broker and the access mechanisms used is implementation-dependent.[1].

1.3: Service requester
The service requester, also known as a service client, discovers services and then uses them as part of its operation.

A service requester uses services provided by service providers. Using an agreed-upon protocol, the requester can find the required information about services using a broker (or this information can be obtained in another way). After the service requester has the necessary details of the service, it can bind or connect to the service and invoke operations on it. The binding is usually static, but the possibility of dynamically discovering the service details from a service broker and configuring the client accordingly makes dynamic binding possible.[1]

To summarize, a complete web service is, therefore, any service that:

- Is available over the Internet or private (intranet) networks
- Uses a standardized XML messaging system
- Is not tied to any one operating system or programming language
- Is self-describing via a common XML grammar
- Is discoverable via a simple find mechanism.[3].

2. BENEFITS OF WEB SERVICES

2.1: Exposing the Existing Function on the Network
A web service is a unit of managed code that can be remotely invoked using HTTP. That is, it can be activated using HTTP requests. Web services allow you to expose the functionality of your existing code over the network. Once it is exposed on the network, other applications can use the functionality of your program.[3].

2.2: Interoperability
Web services allow various applications to talk to each other and share data and services among themselves. Other applications can also use the web services. For example, a VB or .NET application can talk to Java web services and vice versa. Web services are used to make the application platform and technology independent.[3].

2.3: Standardized Protocol
Web services use standardized industry standard protocol for the communication. All the four layers (Service Transport, XML Messaging, Service Description, and Service Discovery layers) use well-defined protocols in the web services protocol stack. This standardization of protocol stack gives the business many advantages such as a wide range of choices, reduction in the cost due to competition, and increase in the quality.[3].

2.4: Low Cost Communication
Web services use SOAP over HTTP protocol, so you can use your existing low-cost internet for implementing web

services. This solution is much less costly compared to proprietary solutions like EDI/B2B. Besides SOAP over HTTP, web services can also be implemented on other reliable transport mechanisms like FTP.[3].

3. TECHNOLOGIES

Over the past two years, three primary technologies have emerged as worldwide standards that make up the core of today's web services technology. These technologies are:[2].

3.1: Simple Object Access Protocol (SOAP)

SOAP provides a standard packaging structure for transporting XML documents over a variety of standard Internet technologies, including SMTP, HTTP, and FTP. It also defines encoding and binding standards for encoding non-XML RPC invocations in XML for transport. SOAP provides a simple structure for doing RPC: document exchange. By having a standard transport mechanism, heterogeneous clients and servers can suddenly become interoperable. .NET clients can invoke EJBs exposed through SOAP, and Java clients can invoke .NET Components exposed through SOAP. [2].

3.2: Web Service Description Language (WSDL)

WSDL is an XML technology that describes the interface of a web service in a standardized way. WSDL standardizes how a web service represents the input and output parameters of an invocation externally, the function's structure, the nature of the invocation (in only, in/out, etc.), and the service's protocol binding. WSDL allows disparate clients to automatically understand how to interact with a web service. [2].

3.3: Universal Description, Discovery, and Integration (UDDI)

UDDI provides a worldwide registry of web services for advertisement, discovery, and integration purposes. Business analysts and technologists use UDDI to discover available web services by searching for names, identifiers, categories, or the specifications implemented by the web service. UDDI provides a structure for representing businesses, business relationships, web services, specification metadata, and web service access points. [2].

4. The Role of WSDL, SOAP, and Java/XML Mapping in SOA

4.1: The Role of WSDL in SOA

WSDL is the interface definition language (IDL) that defines the interactions among SOA components. It provides a standard language for describing how to communicate with a component. Without a standard IDL, you must resort to ad hoc documentation to communicate the interfaces for your SOA components.[4].

Figure 4.1 shows the role of WSDL for SOA Integration as described in this book. The figure provides a UML object diagram depicting the relationship of WSDL in an SOA Integration setting. First, notice the Web Services Platform subsystem where deployment takes place. The top-level class depicted in that subsystem is Service Deployment. Each instance of Service Deployment corresponds to a Web service that is deployed on this platform. Next, notice that Service Deployment contains both an operation (taken from the WSDL interface description) and a Java method. In this manner, you see that a Web service deployment defines a relationship between a WSDL interface description and a Java implementation of that description. More specifically, a Web service deployment defines relationships between individual operations in a WSDL and the Java methods that implement them in that particular deployment.[4].

Figure 4.1 shows that a WSDL interface description contains a types instance (i.e., the wsdl:types element). As you know, this is the top-level element within the wsdl:definitions element that describes the XML Schema types used in the WSDL. Here, you can also see that this particular WSDL's types instance contains the schema Orders.xsd that is part of the XML Schema Library in an Enterprise System. It also incorporates the schema Faults.xsd from the Web Services Infrastructure subsystem.[4].

The Enterprise System in Figure 4.1 could be the OMS described in which case, its schema library would include the Orders.xsd schema described there. Other "infrastructure" libraries—like standard schema for fault messages—are also envisioned as part of the SOA framework deployed by an enterprise.

So, as described here, one role of WSDL in the SOA framework is to assemble standard XML types into operations that describe Web services. Another role is as a participant in a Web service's deployment.[4].

The preceding section discusses how SOA requires an interface definition language (i.e., WSDL) to describe, in a standard way, how to invoke a Web service component. The WSDL describes an abstract interface (i.e., wsdl:portType and wsdl:operation), and a concrete binding of that interface (i.e., wsdl:binding). As shown in Example 4–2, a WSDL operation is described in terms of abstract messages (i.e., wsdl:message elements). SOAP provides a concrete implementation, or binding for the wsdl:message elements, thereby defining the XML structure of the messages exchanged among SOA components in a standard manner. In this paper, SOAP means SOAP Version 1.1 unless a specific reference is made to SOAP Version 1.2. [4].

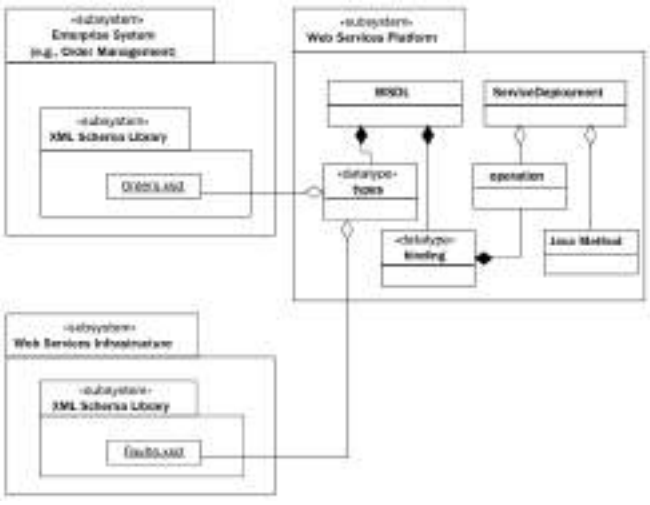


Figure 4.1: The role of WSDL in SOA integration.

4.2: The Role of SOAP in SOA

REST proponents also argue that REST is simpler than SOAP because it does not concern itself with the semantics of SOAP nodes. Along with its envelope structure, SOAP includes a processing model composed of SOAP nodes that transmit and receive SOAP messages, and may relay them to other SOAP nodes. SOAP even goes so far as to prescribe header attributes such as env:mustUnderstand that are used to indicate whether processing of a SOAP header block is mandatory or optional. Complexity creeps in this way because if the ultimate receiver node of a SOAP message cannot process a header block that is marked as env:mustUnderstand, it must reply with a SOAP fault. And the SOAP specification describes a standard envelope structure for SOAP faults. [4].

SOAP node semantics and the associated header processing attributes give the REST advocates most of their ammunition for declaring that SOAP is too complicated. I must admit that I have some sympathy for this point of view. However, at this point, the lack of a standard IDL for REST makes it impossible to work with as a standard for enterprise SOA. [4].

4.3: The Role of Java/XML Mapping in SOA

Java/XML mapping for SOA is accomplished by defining and implementing type mappings. A type mapping is simply a relationship between a Java class and an XML Schema type—for example, corp:AddressType and samples.Address. A type mapping is implemented by a serializer and a deserializer. The serializer converts instances of the Java class into XML instances that conform to the schema. The deserializer does the reverse. [4].

When doing "Start from WSDL and Java" development, a large portion of the design process involves defining the type mappings and serializers. For example, suppose you have a Java method such as:

public void updateAddress(String custId, Address addr)

And suppose the corporate standard schema for address is a complex type: corp:AddressType. Then, the WSDL that is deployed to describe the Web service for updateAddress needs to include corp:AddressType as a message part. But furthermore, the serialization subsystem on the platform where the Web service is deployed must be able to access the deserializer for the type mapping . When a SOAP request for the Web service arrives, the deserializer is used to convert the SOAP part to the Java method parameter. This process is illustrated in Figure 4.2. [4].

As discussed in Section 4.3, the dispatching of this SOAP message is based on the wrapper element—custinfo:updateAddress. It gets mapped to the updateAddress method as shown. Below this wrapper element are the two message parts8 —custinfo:custId and custinfo:address. These are mapped to the parameters custId (String) and addr (Address), respectively. This mapping, of the message parts to the method parameters, is not defined in the WSDL. The WSDL contains no information about the underlying Java implementation of the Web service. This property of the WSDL is consistent with the separation of concerns concept discussed in Section 4.2. After all, the consumer of the Web service shouldn't have to be concerned with such implementation details. All the consumer needs is the information necessary to construct the SOAP message and send it to the appropriate URL. [4].

So, the type mappings that link the SOAP/WSDL to the Java implementation are not defined in the WSDL, but rather are part of the internal—platform-specific—deployment information associated with the Web service. In the JWS model, these type mappings are defined by the JAXB standard mapping as customized by any annotations.

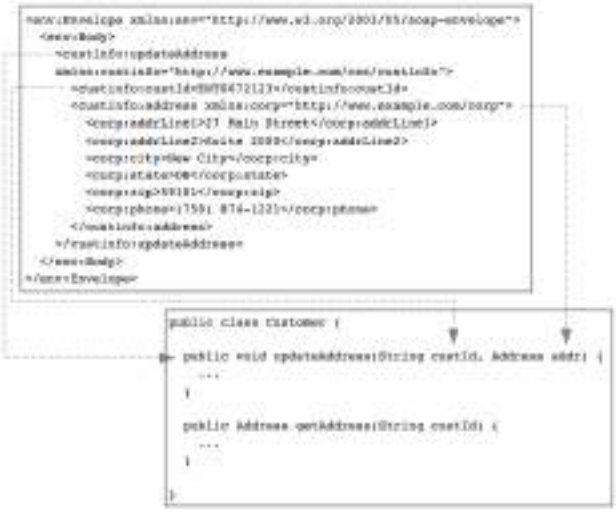


Figure 4.2: SOAP parts map to Java method parameters.

At this point, we just want to point out that the type mapping process is outside the scope of the WSDL and is a platform-specific issue. Furthermore, you should understand that being able to implement flexible type mappings is a key to the "Start from WSDL and Java" development model needed for SOA. As illustrated in Figure 4.2, the key to being able to deploy the updateAddress() method as a Web service with the desired WSDL is to be able to implement the type mapping . [4].

CONCLUSION

In this paper we tried to bridge the technologies of web Services and SOA. Critical Web services infrastructures will be covered, such as WSDL, SOAP, UDDI, Discovery, Composition, Registry, and Web services invocation and relationship binding. How Web and SOA can benefit with each other will also be explored and also about the Role of WSDL, SOAP, and Java/XML Mapping in SOA. Finally, the presenter will depict technologies and development challenges and directions in the field of SOA and Web services. [5].

REFERENCE

[1].http://www.redbooks.ibm.com/redpapers/pdfs/redp4884.pdf

[2].http://gsl.mit.edu/media/programs/south-africa-summer-2015/materials/o'reilly_-_java_web_services.pdf

[3].http://www.tutorialspoint.com/webservices/webservices_tutorial.pdf

[4]. Mark D. Hansen, *SOA Using Java Web Services*, Prentice Hall,2007.

[5].http://www.intersystems.com/assets/SOA_WP-be1cf67974f8d09552742ccdbba0792d.pdf

Karthik.V studying M-Tech (Computer Network Engineering) in New Horizon college of Engineering, marathalli. He completed BE (Information Science Engineering) degree in VTU University in East Point college of Engineering and Technology. His area of interest includes Data Mining, SOA, and Computer Networks.

R.JAYA is working as a Sr.Assistant Professor in the department of CSE at New Horizon college of Engg.She has 13 years of teaching experience. She has done bachelor and master degree in engineering under Anna University. She is doing Ph.D under VTU. Her research interest includes data mining, expert systems,SOA and security.

Study of IoT: Understanding IoT Architecture, Applications, Issues and Challenges

Soumyalatha, Shruti G Hegde
Assistant Professor, Department of Computer Science and Engineering
Brindavan College of Engineering, Bangalore-63, India
soumyanaveen.u@gmail.com, shruti601@gmail.com

ABSTRACT-The Internet of Things (IoT) is an emerging and challenging field for researchers. IoT is a network of general objects which are embedded with technologies that helps to communicate and interact within themselves and external environment. This in-turn provides intelligence to the objects to make people life comfortable. In this paper we discussed IoT, its architecture, Relationship between wireless sensor networks (WSN) and IoT. Further we explained different applications of IoT, Current IoT tools for users, IoT advantage, disadvantage and challenges. In this paper we also proposed an idea of using IoT in Indian agriculture domain.

Keywords—Internet of things, wireless sensor networks, sensor, architecture, Smart city, IOT Tools, Smart agriculture.

1. INTRODUCTION

Evolution of internet began by connecting computers. Later many computers were connected together which created World Wide Web. Then mobile devices were able to connect to the internet which leads to mobile-Internet technique. People started using the internet via social networks. Finally the idea of connecting daily objects to the internet was proposed, which lead to the Internet of Things technology [1].

First time the term "Internet of Things: word was used by Kevin Ashton in a presentation during 1998. [2]. He has mentioned "The Internet of Things has the potential to change the world, just as the Internet did May be even more so". Later during 2001, MIT AutoID Lab center presented their view on IoT. Then during 2005, this is formally recognized by the International Telecommunication Union (ITU).

IoT creates a world where all the objects (also called smart objects) around us are connected to the Internet and communicate with each other with minimum human intervention. The ultimate goal is to create 'a better world for human beings', where objects around us know what we like, what we want, and what we need and act accordingly without explicit instructions.

Current research on Internet of Things (IoT) mainly focuses on how to enable general objects to see, hear, and smell the physical world for themselves, and make them connected to share the observations. In that sense, monitoring and decision making can be moved from the human side to the machine side.

Since IoT is considered as the networked connection of physical objects or devices. One of the definition of IoT by a researcher is [3]"An open and comprehensive network of intelligent objects that have the capacity to auto-organize, share information, data and resources, reacting and acting in face of situations and changes in the environment".

With the help of the communication technologies such as wireless sensor networks (WSN) and Radio frequency identification (RFID), sharing of information takes place [4]. So in general we can say IoT allows people and things to be connected Anytime, Anyplace, with anything and anyone using any network and any service as shown in fig 1.

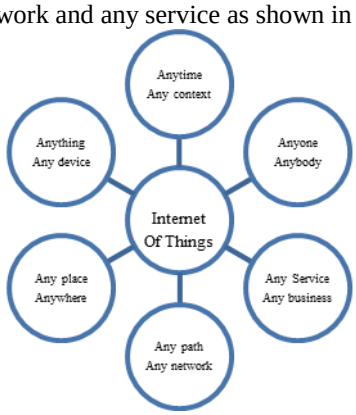


Fig.1 Definition of IoT.

Following table shows the relationship between sensor network and IoT. [5,6]

TABLE I: Relationship between Sensor networks and IoT

Senor Network	IoT
Sensor is a device, which collects data.	Devices or objects are made smart objects, which is connected to the internet.
Sensors are part of IoT.	IoT is combination of sensors, network and people
Data is collected using sensors.	Data is collected , later it is processed and decisions are taken
Especially sensors are used to monitor space, objects and human beings.	Here daily life objects are made intelligent, which alerts when something is wrong.

According to, Cisco Internet Business Solutions Group (IBSG) study, in 2003, there were 500 million devices[7] connected to the Internet and approximately 6.3 billion people were living. Explosive growth of smart phones and tablet PCs brought the number of devices connected to the Internet to 12.5 billion in 2010, while the world's human

population increased to 6.8 billion. It also predicts there will be 25 billion devices connected to the Internet by 2015 and 50 billion by 2020.

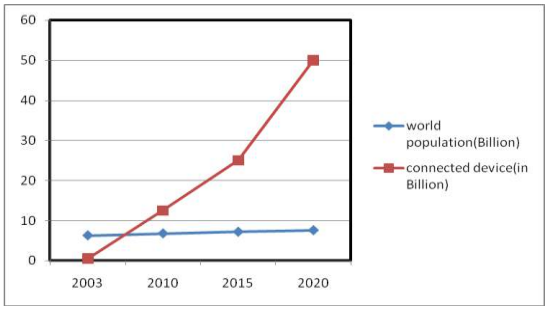


Fig.2: Year Vs world population and connected devices.

2. ARCHITECTURE OF IOT

Architecture of IoT [8] is broadly classified into 4 layers.

2.1 Sensor Layer

This is lowest layer of IOT Architecture, which consists of sensor networks, embedded systems, RFID tags and readers or other soft sensors which are different forms of sensors deployed in the field. Each of these sensors has identification and information storage (e.g. RFID tags), information collection (e.g. sensor networks), etc [9]

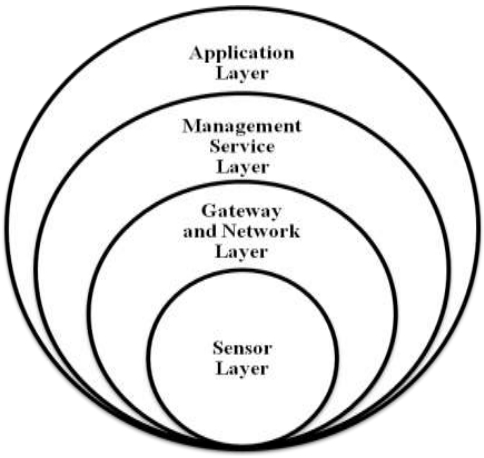


Fig. 3 Layered Architecture of IOT

2.2

ccess Gateway and Network Layer

This layer is responsible for transferring the information collected by sensors to the next layer. It should support scalable, flexible, standards universal protocol for transferring data from heterogeneous devices (Different types of sensor nodes). This Layer should have high performance and robust network. It should also support multiple organizations to communicate independently.

2.3 Management Service Layer

This layer acts as an interface between the Gateway - Network layer and the application layer; in bidirectional mode. It is responsible for device management and information management and responsible for capturing large amount of the raw data and extracting relevant information from the stored data as well from the real time data. Security and privacy of the data should be ensured.

2.4 Application Layer

This is the top most layer of IoT which provides a user interface to access various applications to different users. The applications can be used in various sectors like transportation, health care, agriculture, supply chain, government, retail etc.

3 IOT APPLICATION DOMAINS

This technology has a lot of applications in various fields. Following are some possible areas where we can leverage the power of the Internet of Things (IoT) to solve day-to-day problems. However, it can be put to many more uses.

3.1 Smart society

3.1.1 Smart home

Now a day homes and offices use IoT technologies. Various electronic gadgets and HVAC systems such as lights, fans microwave ovens, refrigerators, heaters and air conditioners are embedded with sensors and actuators to utilize the energy sufficiently, monitor and control amount of heating, cooling and level of light, ,room light sense the presence of human beings and turn on when you enter , when fire or smoke detected at home, wireless smoke and carbon monoxide sensors sound alarms and also alert by phone or email and adds more comfort in life, which in turn minimize the cost and increases energy saving. [10]

3.1.2 Smart home automation

The IoT can be used to remotely control and program the appliances in your home **[**11**]**. It can be useful in detecting and avoiding thefts.

3.1.3 Smart City

On a broader scale, IoT technologies can be employed to make cities more efficient. The goal of smart cities is to leverage the IoT to improve the lives of citizens by improving traffic control, monitoring the availability of parking spaces, evaluating air quality and even providing notification when trash containers are full.

3.1.4 Smart Traffic

Currently the traffic management is a bigger issue in the metropolitan cities. Managing them manually has become almost impossible. This problem can be overcome by implementing IoT for traffic management. This smart traffic monitoring uses sensors to collect raw traffic data, which provides traffic update to driver, which helps him to make the decision for travelling better route. This also helps user to book a cab without phone call or pick up location and also shows cabs close and also their movement in real time.

3.1.5 Smart parking Sensors will be placed in parking slots to know whether parking slot is available or not. The drivers park their vehicle looking into the application which provides the details of nearest parking slots available, parking cost based on the data collected and analyzed by the smart sensors which helps them to save time and fuel.

3.1.6 Smart waste Management

A trash bin embedded with sensors which are capable to analyze and alert the authorities when it is full and need to be emptied.

3.1.7 Smart Street light Sensors which can analyze the context such as time, season, weather conditions will be embedded within street lights which automatically turns light on or off and sets the dimming levels of individual or group of lights based on the context.

3.1.8 Smart water supply Smart cities must monitor water supply to ensure that there is adequate access of water for resident and business need. Wireless Sensor helps to monitor their water piping systems more accurately and discovers water leakage and alerts them about the water loss which in turn saves money and natural resource too.

3.2 Smart Environment

A very important application of IoT is detecting pollution and natural calamities. We can monitor the emissions from factories and vehicles to minimize air pollution. We can track the release of harmful chemicals and waste in rivers and the sea, thereby arresting water pollution. We can also keep tabs on the quality of water being supplied for drinking. We can send warnings of earthquakes and tsunamis by detecting tremors. We can keep the water level of rivers and dams under surveillance to be alert in case of floods. The detection of forest fire is also possible with this technology.

3.2.1 Air Quality Monitoring

By embedding sensors which collects context information such as amount of carbon monoxide (CO), nitrogen dioxide (NO2) in the air, sound levels, temperature, humidity levels in the environment. This provides continuous information about the context, which helps to take the precaution if it exceeds the normal level.

3.2.2 Smart Water Quality Monitoring

Sensors which can detect context such as water quality, water flow, speed, temperature, water pollution, contents of water are placed or flowed in the water. This helps in real-time analysis and management of the water resources available for use.

3.2.3 Smart sewage water management

Embedded sensors in sewage tank, helps to control the overflow of the waste water flowing into; by continuously providing the information about the level of waste water stored. By these data, the maintenance people can schedule the water treatment process to avoid the overflow of sewage.

3.2.4 Natural Disaster Monitoring

Natural disasters such as earth quake, landslides, forest fire, volcanoes, flood, etc. can be predicted by using wireless detection sensors. These detections intimate the respective authorities to take the precautions before the disaster occurs.

3.3 Agriculture

3.3.1 Smart Farming

Context information such as current temperature, soil moisture conditions, leaf wetness, and solar radiation is collected and analyzed by the sensors, which in turn informs the owner regarding the requirements of water, pesticides, manure, fertilizer or treatment for the infected plants.

3.4 Healthcare

3.4.1 Health tracking

The IoT is used in healthcare domain to improve the quality of human life by assisting basic tasks that humans must perform through application. Sensors can be placed on health monitoring equipment used by patients. The information collected by these sensors is made available on the Internet to doctors, family members and other interested parties in order to improve treatment and responsiveness. Additionally, IoT devices can be used to monitor a patient's current medicines and evaluate the risk of new medications in terms of allergic reactions and adverse interactions. With the use of sensors and the technology stated above we can track the person's body temperature, heart beat rate, blood pressure, etc. In case of emergency, the individual and their personal doctor will be notified with all the data collected by the sensors. This system will be very useful to senior citizens and disabled people who live independently.

3.4.2 Pharmaceutical products

Safety of pharmaceutical product is of utmost importance to prevent the health of patients. Attaching smart labels to drugs, and monitoring their status with sensors has benefits like maintaining the storing conditions, expiry of drugs which will prevent the transporting the expired medicines to the patients [12].

3.4 Food sustainability

Packed Food that we eat has to go through various stages of food cycle such as production, harvesting, transportation and distribution. Sensors are used to detect context like temperature, humidity, light, heat etc. which precisely notifies the variation and intimates the concerned persons to prevent the food from spoiling.

3.5 Supply-chains

Internet of Things monitors every stage of supply chain from purchasing of raw materials from the suppliers by the manufactures, production, distribution, storage, product sales and after sales services. This will help to maintain the stock required for continuous sale, which in turn results in customer satisfaction and increased sales [12]. According to Cisco's economic analysis, IoT will generate $1.9 trillion from supply chain and logistics over the next decade[13]. By using this we can also diagnose if the machines require repair and maintenance.

Even the Indian Government has planned to develop 100 smart cities which cover some of the above mentioned IoT applications.

4 IOT TOOLS FOR USERS

IOT can also function as a tool that saves people's money and time. Some of the tools used are:

- HAPIfork[14]—Is an electronic fork to monitor and track the eating habits of user. When the user is eating too fast it alerts with the help of indicator lights and gentle vibrations.

- HeatWatch[15]- is a cattle monitoring solution that records the activities of animal which in turn helps farmer to breed more cows.
- Smart Traffic ParkSightis*[16]:* Is a parking management application to guide the driver regarding the availability of parking slots.
- SmartBelly[17] : is a smart waste management solution to alert authorities when the trash bin is full .
- Airqualityegg[18]: is an application which notifies the amount of air pollution.
- Aircasting[19] is a platform for sharing health (heart and breathing rate, pressure, etc.) and environmental data(temperature, humidity, sound level, air contents).
- G.Dontflush[20]- is an application to intimate peoples to reduce the usage of water when the sewage tank is full.
- Insightrobotics[21]- is a forest fire detection system.

Some other tools are bumblebee , Floating Sensor Network and Intelligentriver , MyVessylCup , Smart Tooth brush, Smart propane tank, Glucose monitoring, Smart Washing machines, Hydroponic System, Smart sprinkler Controller, Smart home security, Smart lightening, Smart a/c, Blood pressure monitor, Smart weather station, Smart slow cooker, Smart bike, Smart garbage cans, Smart gardening.

5 THE ADVANTAGES AND DISADVANTAGES OF IOT

5.1 Advantage

5.1.1 Communication

Since IoT has communication between devices, in which physical devices are able to stay connected and hence the total transparency is available with lesser inefficiencies and greater quality.

5.1.2 Automation and Control

Without human involvement, machines are automating and controlling vast amount of information, which leads faster and timely output.

5.1.3 Monitoring saves money and time

Since IOT uses smart sensors to monitor various aspects in our daily life for various applications which saves money and time.

Fig.4 Advantages and Disadvantages of IoT

5.1.4 Better Quality of Life

IoT based applications increases comfort and better management in our daily life; thereby improving the quality of life.

5.1.5 New business opportunities

Creates new business for IoT technology, hence increases economic growth and new jobs .

5.1.6 Better Environment

Saves natural resources and trees and helps in creating a smart greener and sustainable planet.

5.2 Disadvantages

5.2.1Compatibility

As devices from different manufacturers will be interconnected in IoT, presently , there is no international standard of compatibility for the tagging and monitoring equipment.

5.2.1 Complexity

The IoT is a diverse and complex network. Any failure or bugs in the software or hardware will have serious consequences. Even power failure can cause a lot of inconvenience.

5.2.2.Privacy/Security

IoT has involvement of multiple devices and technologies and multiple companies will be monitoring it. Since lot of data related to the context will be transmitted by the smart sensors, there is a high risk of losing private data.

5.2.3 Lesser employment of menial staff:

With the advent of technology, daily activities are getting automated by using IoT with less human intervention, which in turn causes fewer requirements of human resources. This causes unemployment issue in the society.

5.2.4 Technology Takes Control of Life

Our lives will be increasingly controlled by technology, and will be dependent on it. The younger generation is already addicted to technology for every little thing. With IoT, this dependency will spread amongst generations and in daily routines of users. We have to decide how much of our daily lives are we willing to mechanize and be controlled by technology.

5.3 Challenges and issues

5.3.1 Scalability

Spontaneously various new smart objects or devices are getting connected to the network. So IoT should be capable to solve the issues such as addressing, information management and service management and also should support both small-scale and large-scale environments**.**

5.3.2 Self configuration

IoT objects should be programmed for self configuration to suit particular environment without manual configuration by the user.

5.3.3 Interoperability (Devices heterogeneity)

In IoT many smart objects are connected and each smart object has its own information collection capability, processing and communication capability. For communication and cooperation between the smart objects of different types; they should have common communication standard.

5.3.4 Software complexity

Since software systems in smart objects work with minimal resources, there is a need for software infrastructure to

support the network and requires a server on the back ground to manage and support smart objects of the network.

5.3.5 Storage Volume

Based on the scenario and context, smart objects collect either small amount of data or huge volume of data. So based on amount of data, storage has to be allocated.

5.3.6 Data interpretation

It's very important to interpret the context, that sensor has to sense. Context has important role for generating useful information and to draw a conclusion from the data sent by the sensors.

5.3.7 Security and personal privacy

In IoT, network formed by smart objects via internet, so providing security and privacy is a big challenge. In IoT, sometimes user prevent other user to access some particular information at certain time or preventing some communication or some transaction to protect secrete information from competitors. So handling all this situation is a big challenge.

5.3.8 Fault tolerance

In IoT, smart objects or devices are dynamic and rapidly context may change. But still network has to function properly automatically, to adapt to the changed conditions. So IoT has to be structured for fault tolerance and robust [22].

5.3.9 Ubiquitous data exchange through wireless technologies:

Issues such as availability, network delays, and congestion etc. of wireless technologies; which is used for communication of smart devices are big challenge.

5.3.10 Energy-optimized solutions

Network consists of many interconnected devices; which requires high energy to keep the network active. So energy optimization is the major aspect in IoT[23].

6 PROPOSED IDEA FOR AGRICULTURE

Traditionally in India farmer usually follows the following major phases for agriculture. They are Crop Selection, Land Preparation, Seed Selection, Seed Sowing,, Irrigation, Crop Growth, Fertilizing and Harvesting. They were completely dependent on man power for all the above mentioned work. Now a day's finding the man power has become tedious task and even the profit is reduced. This has lead to the decrease the number of peoples completely involving in agriculture. A solution is necessary so that the owner being least dependent on others; can manage his own field and gain the best returns from what he grows.

6.1 How technology can help?

By using IoT in different phases of agriculture is a solution to the above mentioned problems. In this multiple sensors are embedded in the fields, which collects real-time information regarding weather, temperature, humidity, rain fall, soil moisture, soil composition, wind speed, wind direction, soil temperature, leaf wetness, air quality, predicting pest, crops , water level, which in turn, this predictive statistical data provides information to the owner to make smarter decisions. This information helps the owner/farmer in crop selection, crop monitoring, crop maturity, crop yield, spreading of seeds, fertilizers, pesticides, soil erosion, crop yield, diagnosis of crop diseases.

Further IoT sensors enabled machineries could further improve yields. These leads to less man power, increased productivity and profit.

6.2 Present Tools for agriculture

Cropln technology[24] has developed a mobile application which takes input related to farm and helps in efficient maintaining and ensuring the crop quality in short duration.

Precisionhawk**[**25] provides an application for the field of agriculture, which asses plant health, allocates water, detects weeds and monitor seasons.

Sensefly[26] provides eBee presision agriculture drone. This can capture high quality images of fieild, which in turn helps the farmer to take decisions.

Based on the survey, there are only few tools or applications are available for agriculture. So we propose the idea of embedding multiple sensors in the field which collects the information of crop health and detects the pests if any and this sensors collectively send the data through the gateway, which will be analyzed and results will be passed on to the smart phones of the farmers via application. Based on these data, farmer can accurately know the affected area and can take right decisions at right time.

Initially creating awareness regarding this new technology is very important. This can be done by providing free demo's and training to the farmers on how to use and the advantages of using this application and also providing them with free samples sensor kits. The opinion of usage and results of the application can be collected and analyzed for further improvement of application.

7 CONCLUSION

In this survey paper, we analyzed IoT domain by considering its architecture and applications and advantage and disadvantage. We observed that still IoT is not much used in the field of agriculture. So we find its very much necessary to improve the applications of IoT in this field and educate the same to the agriculturist, this will in turn reduces the dependency on man power and also will improve the yield which leads increase in the economy.

REFERENCES

[1] Charith Perera, Chi Harold Liu, Srimal Jayawardena *"The Emerging Internet of Things marketplace From an Industrial Perspective: A Survey"*,IEEE transactions on emerging topics in computing", 31 Jan 2015

[2] Jayavardhana Gubbia, Rajkumar Buyyab, Slaven Marusic a, Marimuthu Palaniswami, *" Internet of Things (IoT): A vision, architectural elements, and future directions",* Future Generation Computer Systems, Elsevier,2013.

[3] Somayya Madakam, R. Ramaswamy, Siddharth Tripathi

Internet of Things (IoT): A Literature Review Journal of Computer and Communications, 2015, 3, 164-173
Published Online May 2015 in SciRes.

[4] Tuhin Borgohain, Uday Kumar, Sugata Sanyal, *"Survey of Security and Privacy Issues of Internet of Things"* Int. J. Advanced Networking and Applications *Volume: 6* Issue: 4 Pages: 2372-2378 (2015) ISSN: 0975-0290

[5] Delphine Christin, Andreas Reinhardt, Parag S. Mogre, Ralf Steinmetz, *"Wireless Sensor Networks and the Internet of Things: Selected Challenges"*, pp 31-33 .

[6] Alcaraz, P. Najera, J. Lopez, and R. Roman, *"Wireless sensor networks and the internet of things: Do we need a complete integration?"* in 1st International Workshop on the Security of the Internet of Things (SecIoT'10), 2010.

[7] Joseph Bradley, Joel barbler, Doug Handler , *"Embracing the internet of everything to capture your share $14.4 Trillion"*, , Cisco, 2013

[8] ZHANG Ying-conga, YU Jingb, " A Study on the Fire IOT Development Strategy", Int. J. of Safety and Security Eng., *Vol.* 4, No. 2 (2014) 135–142

[9] Debasis Bandyopadhyay , Jaydip Sen *," Internet of Things - Applications and Challenges in Technology and Standardization",*Wireless Pers Commum,9 April 2011.

[10] Andrew Whitmore & Anurag Agarwal & Li Da Xu, *"The Internet of Things—A survey of topics and trends",*InfSyst Front ,2015.

[11] J. Sathish Kumar, Dhiren R. Patel, *"A Survey on Internet of Things: Security and Privacy Issues",* International Journal of Computer Applications (0975 – 8887) *Volume 90 – No 11*, March 2014

[12] Shashank Agrawal, Dario Vieira, *"A survey on Internet of Things"*,may 2013.

[13] James macaulay, dr. markus kückelhaus, *"Internet of things in logistics",* a collaborative report by dhl and cisco on implications and use cases for the logistics industry, 2015.

[14] https://www.hapi.com/product/hapifork

[15] http://animalsciences.missouri.edu/reprod/Estrus/HDcattleshp/sld008.html

[16]http://aramidesign.com/wp-content/uploads/2014/03/Streetline-DS-ParkSight.pdf

[17] http://cse-cleansolution.com/indexx/en/products/big-belly2/smartbelly

[18] http://airqualityegg.com/

[19] http://aircasting.org/

[20]dontflush.me/

[21] http://www.insightrobotics.com/

[22]Prajakta Pande , Anand R. Padwalkar , *"Internet of Things –A Future of Internet: A Survey", Volume 2,* Issue 2, February 2015.

[23] Ashvini Balte, Asmita Kashid, Balaji Patil , *"Security Issues in Internet of Things (IoT): A Survey", Volume* 5, Issue 4, 2015,

[24] http://cropin.co.in/products/

[25] http://www.precisionhawk.com/

[26] https://www.sensefly.com/

Mrs.Soumyalatha Received BE and M.Tech degree from VTU, Belgaum. Currently working as an assistant professor in the department of CSE at Brindavan college of Engineering, Bangalore. Research interests includes Big data, Internet of Things and Wireless networks

Ms.Shruti G Hegde Received BE and M.Tech degree from VTU, Belgaum. Currently working as an assistant professor in the department of CSE at Brindavan college of Engineering, Bangalore. Research interests includes Big data, Internet of Things and Wireless networks.

A Novel approach to design a Smart bin using through IoT

Sruthi K V, Manjunath K N
Department of Electronics and Communication Engineering, Jain University, Bangalore
sruthisony11@gmail.com, manju23687@gmail.com

ABSTRACT- India is the second largest population country, so collecting and disposal of waste plays a vital role in day today life. This work proposes a clean city concept using Smart bin and its application though a proper interaction and networking with one to many devices. The paper presents a novel waste collection technique and interaction through a Smart bin which is developed using ARM LPC 2148 with ultra-sonic sensors and pressure sensing resistor .The proposed system also provides the web page interactions to the terminal side with effective data base management and alert system according to the function of the bin. HTML web page is being created to show various levels of every bin located in each ward with various levels of trashes. A GSM transmits an alert to the terminal of the particular bin. RFID were incorporated for the authentication.

Keywords: Force sensing resistor, GSM, HTML, Smart bin, RFID, Ultrasonic sensor.

1. INTRODUCTION

ow a days waste management has become a sensational Nissue. To make SWACHBHARATH great success all the Indian citizens has to work together, it is quite challenging As the population increases the waste which is accumulated in the public bins are overfills and the waste will not carried to space land in time. It will spread diseases and effects the cleanness of the city. There are many solutions were proposed for the waste management concern these are not become successful due to many criteria, control unit has been placed in the dustbin in every public areas. But reality is everywhere the trash bins are overfilling and overflowing so the area is becoming untidy also area will be prone to diseases.

The manual method of tracking out the trash bins are tedious because sometimes a few bins will fill fast and some other are late. A Layman should be go to every bins and need to empty irrespective of the filling levels. It unnecessarily wastes human resource. Filling of waste depends upon not only the persons using the bin but also depends on the special occasions such as festivals, marriage etc. There is a survey carried out by Ranjith Kharvel Annepu, it cited that per data there is 130000 tons of Municipal Solid Waste (MSW) is produced in urban areas of India, which is capable of producing 1751 kcal/kg of energy [2] shows in Table 1. So, in this paper we have integrated analytics and electronics in order to create optimal changes in the conventional methodology of waste collection with the large amount of data that is being produced by the smart bin networks. The movement of waste across the whole city can be tracked and thus can be monitored by a single system efficiently and concretely. This system can prove to be a revolution for the whole urban waste management system of upcoming smart cities

Table 1: statics of MSW in India

Region city	MSW (TPD)	Compostible %	Recyclable %	Inerts %	Moisture %	Cal value MJ/Kg	Cal value Kcal/kg
Metros	51400	50.89	16.28	32.82	46	6.4	1525
Other cities	3723	51.91	19.23	28.86	49	8.7	2084
East India	380	50.41	21.44	28.15	46	9.8	2341
North India	6825	52.38	16.78	30.85	49	6.8	1625
South India	2345	53.41	17.02	29.57	51	7.6	1827
West India	380	50.41	21.44	28.15	46	9.8	2341
Overall Urban India	130000	51.30	17.48	31.21	47	7.3	1751

2. EXISTING SYSTEMS

There are many steps being taken to avoid the overfilling of dustbins. Waste management becoming tougher and tougher these days. Large number of workforce are appointed for cleaning the metropolitan cities. Even though the task has not been reached to expected level.

A system [3] proposed which helps to show the nearby dustbin in a city. Users are voluntary need to log into the webpage to find out the bins. The System [5] introduces the user to put recycling material, users will be gets points by registering with their RFID. In advance to the existing system [5], this design [6] found little flexibilities which gives the high degree of scalability in terms of collecting the waste, each area is divided into polygon for every area a vehicle is provided the sensor data will go to vehicle where there is onboard computer. But this is failed to its cost and efficiency.

A few other system [2] which will alarm the system authority about the fullness they need to pick up waste in the

time scheduled manner but there is no controlling action is taken if they failed to pick up the trash bin in time.

2.1 Comparison of existing systems with the proposed system

By analyzing the Table 2, there are many systems had been developed to control and monitor the Municipal Solid Waste (MSW). The systems were successfully implemented with finding the nearest bin location, plastic sorting, recycling waste collection etc. But they are failed due to cost and precision of various sensors. In the proposed system it is mainly using the Internet of Things(IoT) also the GSM technology, Informing the various filling levels of the bin is become quite easy through the ultrasonic sensor and pressure sensor ,using the RFID identifying the location and authentication can be possible.

3. SYSTEM DESIGN

Taking the considerations of all the possible problems, proposed a smart bin which can sense the various levels of the bin using the ultrasonic sensors and also has the potential to measure the weight using pressure sensor.

The bin will be having various LED indication for various level. Namely for 20%,50%,70%, if the bin is 90% then filled the it will send a message to administer saying that the bin with this ID is about to fill please come and collect the waste. If the person delays it, the bin has been closed automatically and also there is an alarm to the user who dumps again the waste into the filled bins

The proposed bin will be fixed and the bin will be attached with the RFID. The layman with RFID reader only open the bin. So that can avoid the theft of the bin because the system is placed with sensors and controller. The Figure1 shows the basic block diagram of the proposed system.

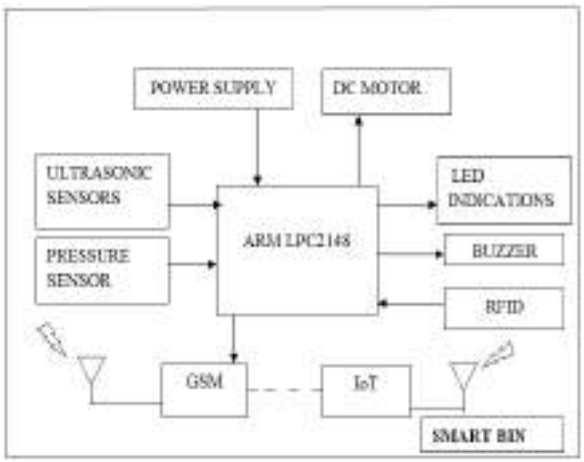


Fig 1: proposed system

Terminal Unit**:** The sensed data is sent to the terminal administrator where the data is stored in the database. The GUI is provided through the HTML where administer has the graphical representation of each bin. Through the routine statistics one can make some essential observation like amount of thrash collected per daily, weekly, and monthly basis which helps to keep more trash bins for the required locations

4. SYSTEM SPECIFICATIONS

4.1 ARM LPC2148 :This controller is 32 bit, flash memory ranging from 32kb to 512kb, memory interface is 128bit it is a very tiny and low power device It is a full speed device with multiple UARTS, SPI, SSP, I2C and on chip SRAM memory ranges from 8kb to 40kb, various 32 bit timer, single or dual 10 bit ADC, PWM channels, also 45 fast GPIO lines with nine edge or level sensitive interrupt this will makes the system which suitable for the high end application[10]

4.2 Ultrasonic sensor: HC-SR04 is the sensor used, which will sense the various levels of the trash in and sends the data to the server HC-SR04 provides ranging between 2cm-300cm it has non-contact distance sensing capabilities, Ranging accuracy up to 3mm, module comprises an ultrasonic transmitter, a receiver and a control circuit[9]

4.3. Radio Frequency Identification (RFID): It being used to authenticate the user who is going to pick up the waste also the people who throws the waste to the bin. It Uniquely identifies an individual item beyond just its product type also it can identify items without direct line-of-sight, it has the potential to identify many items (up to 1,000s) simultaneously, items within a vicinity of between a few centimeters to several meters it can identify.

4.4 force sensing resistor (FSR): it helps to measure the weight, the weight being considered to measure its capacity. PSR are a Polymer Thick Film (PTF) device which exhibits a decrease in resistance with an increase in the force applied to the active surface. Its force sensitivity is optimized for use in human touch control of electronic devices. PSRs are not a load cell or strain gauge, though they have similar properties [9].

4.5 *Global System for Mobile (GSM) sim900*: GSM is used to send the message to the administrator, it is not necessary to have android based phone. It has RS232 is built with Dual Band GSM/GPRS engine- SIM900A, works on frequencies 900/ 1800 MHz. It is well suited for SMS, Voice as well as DATA transfer application in M2M interface. The onboard Regulated Power supply allows you to connect wide range unregulated power supply [9].

4.6. *Hyper Text Markup Language (HTML)*: Hypertext markup language helps to give website where all can log in identify the filed bins. HTML is very dynamic and interactive, and it plays an important role in improving the user experience of the web. It provides more robust structure and organization.HTML helps us to define our own thoughts, layout, and definitions attractive hence creating the user friendly environment.

4.7 *Structured Query Language (SQL)*: is used as the database to store the sensor data it is possible to retrieve and manage data in relational data base.

5.CONCLUSION

This paper presents a new approach for the design of smart bin using LPC2148 and sensors. The work is proposed to provide an effective date base at the terminal side with user GUI interaction. The proposed bin are placed with various level indication for the fullness at the bin through ultrasonic sensors and pressure sensor for the statistics analysis. The system is also provided with a RFID tags for the authenticate and auto control for close and open the bin.

ACKNOWLEDGEMENT

The author would like to thank Karnataka State Council for Science and Technology (KSCST) for funding this proposal (with serial number: 39S_R_Mtech_008). This work is carried out at project laboratory department of electronics and communication engineering, School of engineering and technology, Jain University.

REFERENCES

[1] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi, "Internet of Things for Smart Cities" ,IEEE internet of things journal, vol. 1, no. 1, February 2014.

[2] Shubham Thakker,R.Narayanamoorthi,"Smart and wireless waste management", IEEE sponsored 2nd international conference on innovation in information embedded and communication systems 978-1-4799-6818-3/15/2015 IEEE

[3] Fachmin F olianto,Y ong Sheng Low, Wai Leong Yeow, "Smart Waste Management System", 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) Demo and Video Singapore, 7-9 April 2015

[4] Fachmin Folianto,Young sheng Low,wai Leong Yeow "Smart bin: Smart waste management", ,2015 IEEE 10th international conferenceon intelligent sensors,sensors,sensors network and information processing

[5] Mohd Helmy Abd Wahab, Aeslina Abdul Kadir, Mohd Razali Tomari, Mohamad Hairol Jabbar, "Smart Recycle Bin A Conceptual Approach of Smart Waste Management with Integrated Web based System". IEEE conferences 2014

[6] Vincenzo Catania, Daniela Ventura, "An Approach for Monitoring and Smart Planning of Urban Solid Waste Management Using Smart-M3 Platform", proceeding of the 15th conference of fruct association

[7] Ghenadie COROTINSCHI, Vasile Gheorghiță GĂITAN, "Smart cities become possible thanks to the Internet of Things", 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania

[8] Yann Glouche Paul Couderc, "A Smart Waste Management with Self-Describing objects", SMART 2013: The Second International Conference on Smart Systems, Devices and Technologies

[9] Force sensing resistor integration guide, Ultrasonic Ranging Module HC - SR04, Research design lab, GSM SIM 900Order Code RDL/GSM/13/001/V1.0

[10] User manual UM10139 for LPC214x by PHILIPS, volume 1,Rev.-01 15 august 2005

[11] A. Vakali, L. Angelis, and M. Giatsoglou, "Sensors talk and humans sense towards a reciprocal collective awareness smart city framework," in Communications Workshops (ICC), 2013 IEEE International Conference on, June 2013, pp. 189–193.

[12] "Smart-m3: Free development software downloads at sourceforge.net," November 2013. [Online]. Available: http://sourceforge.net/projects/smart-m3/files/

[13] J. Honkola, H. Laine, R. Brown, and O. Tyrkko, "Smart-m3 information sharing platform," in Computers and Communications (ISCC), 2010 IEEE Symposium on, June 2010, pp. 1041–1046.

[14] Korzun, Lomov, Vanag, Balandin, and Honkola, "Multilingual ontology library generator for smart-m3 information sharing platform," International Journal On Advances in Intelligent Systems, vol. 4, no. 4, pp. 68–81, 2011.

[15] J. Honkola, H. Laine, R. Brown, and I. Oliver, "Cross-domain interoperability: A case study." ser. Lecture Notes in Computer Science, vol. 5764. Springer, 2009, pp. 22–31.

[16] K. Framling, A. Kaustell, I. Oliver, J. Honkola, and J. Nyman, "Sharing building information with smart-m3," International Journal on Advances in Intelligent Systems, vol. 3, no. 4, p. 347357, 2010.

Preparation of Papers for An E-Governance Portal for Tax Payment

Rizwana unnisa, Bhanu k.N
Dept of MCA, Rajarajeshwari College of Engineering,, Bangalore, India
Email: runnisa246@gmail.com , Bhnau.kn@gmail.com

ABSTRACT-Integration Solutions for collecting all the tax at the State level and Central Also the GST bill or Tax related to GST would be considered in this integration. This proposed system will solve most problems and bring transparency and add values to the collections of tax and it helps to find out the defaulters state wise.
Keywords – *checklist, dispatcher, spot visit, tin, vat,*

I. INTRODUCTION

The Spring framework began as one programmer's need to solve a set of problems that Sun had tried to solve through edict in the Enterprise JavaBeans (EJB) Specification. At the time, companies were putting a lot of money and effort into developing web applications. The lack of established industry standards made management nervous, and development, haphazard. As mentioned in [1] and [2].The J2EE specification promised scalability, security, high availability. Enterprise JavaBeans (EJB), as part of the J2EE suite of specifications from Sun, were intended to be as reusable and portable as their non-enterprise counterparts, plain JavaBeans.

As mentioned in [5] Spring provides a POJO (Plain Old Java Object) based configuration environment, a container to manage the instantiation and lifecycle of your POJO components and a framework to help you put into place some established best practices for your applications as mentioned in [3]. The idea behind Spring is that your code should be well-factored, and components kept pristinely. Your components should run with or without a container, and be testable with minimal to no intrusion from outside classes. In essence, your components should have a life outside of the framework. And as complete entities into themselves, these truly modular components should have a loose affiliation with other components but should not be bogged down in these dependencies. Component factoring in this way has become central to object- oriented programming .Fig. 2 explains the architecture in detail.. As mentioned in [4] Hibernate is an Object-Relational Mapping(ORM) solution for JAVA. It is an open source persistent framework created by Gavin King in 2001. It is a powerful, high performance Object-Relational Persistence and Query service for any Java Application. Hibernate sits between traditional Java objects and database server to handle all the works in persisting those objects based on the appropriate O/R mechanisms and patterns.

Fig1:ORM Hibernate

Hibernate is an open source project whose purpose is to make it easy to integrate relational data into Java programs. This is done through the use of XML mapping files, which associate Java classes with database tables. As mentioned in [7] • Hibernate provides basic mapping capabilities. It also includes several other object/relational mapping (ORM) capabilities, including: − An enhanced, object-based SQL variant for retrieving data, known as Hibernate Query Language (HQL). − Automated processes to synchronize objects with their database equivalents. − Built-in database connection pooling, including three open source variants. − Transactional capabilities that can work both stand-alone or with existing Java Transaction API (JTA) implementations. As mentioned in [6]Fig 1 explains in detail .The goal of Hibernate is to allow object-oriented developers to incorporate persistence into their programs with a minimum of effort. It manages the Hibernate Session Factory as a singleton – a small but surprisingly annoying task that must be implemented manually when using Hibernate alone. − It offers a transaction system of its own, which is aspect oriented and thus configurable, either through Spring AOP or Java-5 annotations. Either of these are generally much easier than working with Hibernates transaction API.

Fig 2:MVC architecture

II. SYSTEM ARCHITECTURE

This architecture will describe about the three layers of system which implements the business logic method which is helpful in Spring MVC controller ,the first layer called as web service consists of controller package and Bean package which is used to create the UI .The second layer is called service layer where the request is mapped to database. The last layer is persistence layer which consists of DAO layer where class and interface are been implemented

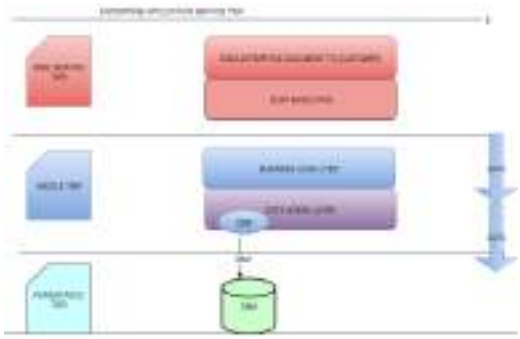


Fig 3.system architecture

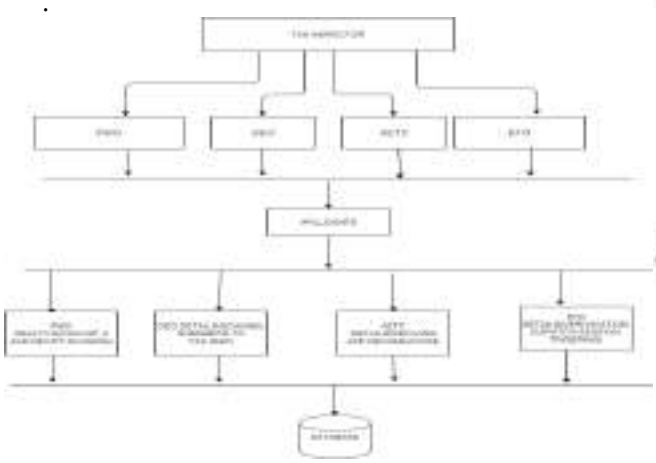


Fig 4.System architecture 2

The same logic of layer has been implemented in this layer too based on tax portal

III. SYSTEM DESIGN

The aim of this portal is to provide complete information about the taxes of different states .There is some problems that need to be solved and bring transparency to the collections of tax.

The above framework shows that dealers/applicant will apply for VAT or GST or TIN through online portal and get acknowledgment receipt along with documents submission. In the next level, the documents are verified by FWO/DEO officer and marked in checklist which is forwarded to tax inspector. Tax inspector conducts the spot visit and once inspection is conducted, AETC officer dispatches the memo tax conduction receipt to dealers

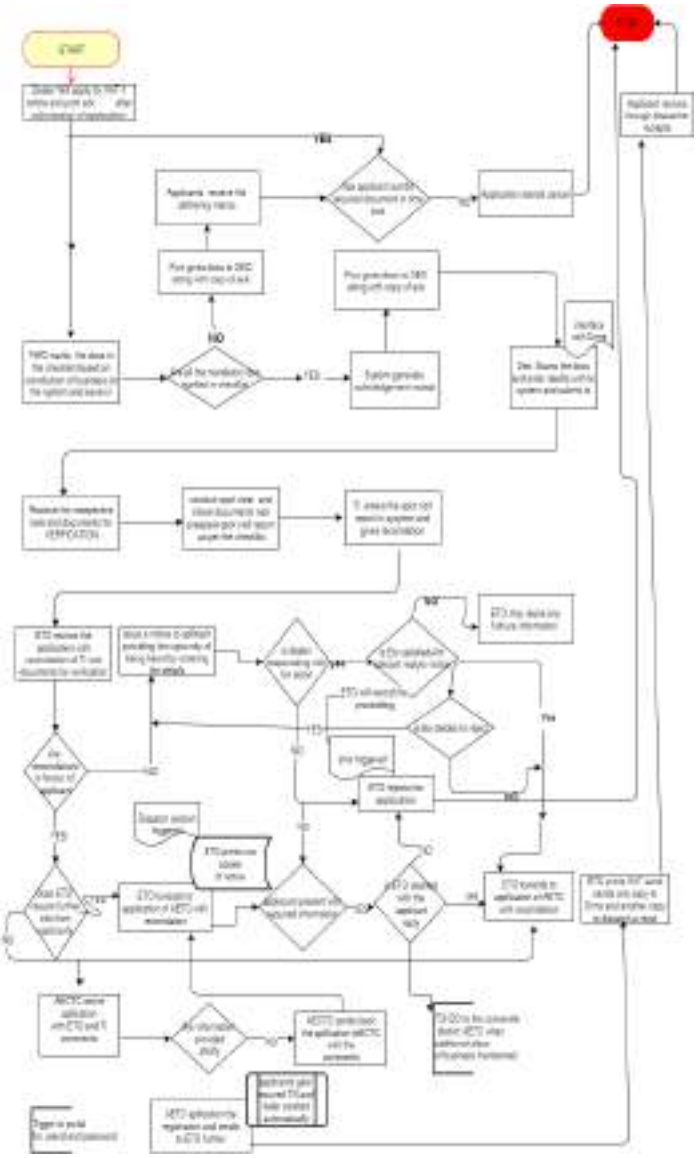


Fig 5. Framework

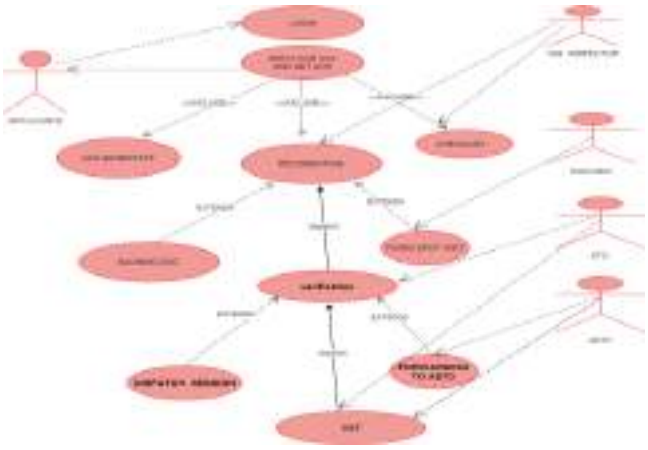


Fig 6. Use case

Use case in Fig.6., explains the actors involved and scenario carried out in automation of this portal.

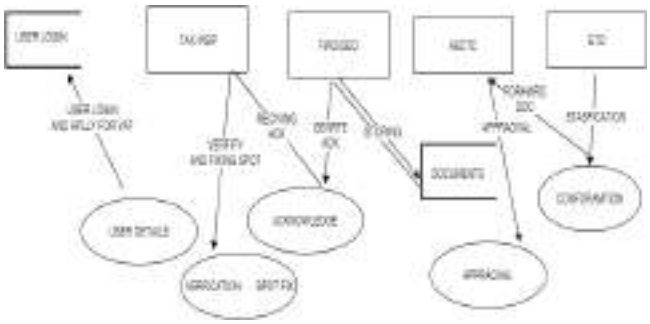


Fig 7. DFD

Data flow diagram in Fig 7. Explains the various tasks carried by administrator to estimate the tax

IV. IMPLEMENTATION

Fig 8. Registration form

In this module the applicants will get registered in the portal to get an acknowledgment from Deo/Fwo officer

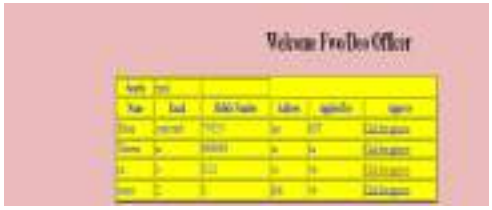


Fig 9.Dashboard of Admin

In this module, the applicants already registered and waiting to get their documents approved from administrator. this module fetches all the registered applicants in the dashboard and allows to get the approved

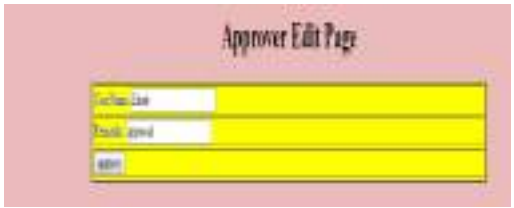


Fig 10 . Approve Module

In this module admin will give approval notify to applicants and the remarks .once the applicants gets the approved acknowledgment their details will be removed from waiting List

Fig 11. Approved page

Fig 12.Applicants details

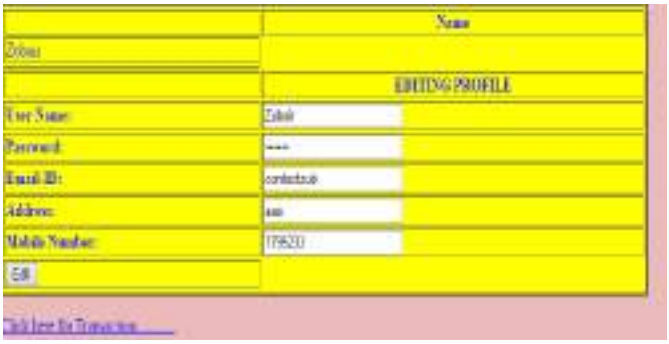


Fig 13.. Profile Module of Applicants

This module allows to applicants to edit their profile only of unread only flied

V. CONCLUSION

The proposed tax integration portal can be used for collecting all kinds of taxes at the state and central level. Also, the GST bill or tax related to GST would be considered in this integration. This proposed system will solve most problems and bring transparency and add values to the collections of tax and it helps to find out the defaulters at the state wise. The aim of our portal is to provide complete information about the taxes of different states .There is some problem to be solved and bring transparency and add values to the collections of tax. The System is a way of disciplined approach towards any task, it's required in any field to have a systematic way of doing the work and preferably

accepted all over the world. A web portal where Central Tax Board and State Tax collection authorities will have the single point where they can upload all the tax collected by state and central .central will have the overlook of the taxes collected by state authorities and give their share accordingly. All the taxes including the new bills would be part of this portal. In our portal, a strong care has been taken to meet the customers need.

ACKNOWLEDGEMENTS

The authors of this paper are thankful to the Management, RRCE, Bangalore and Principal, RRCE, Bangalore for their continuous encouragement and support throughout the work.

REFERENCES

Websites:

[1] [1]. http://spring.io/docs/reference

[2]

[2]. http://www.mkyong.com/tutorials/spring-mvc-tutorials/.

[3] [3]..http://www.capcourse.com/Library/Spring+Hibernate/Spring+Hibernate

Books:

[4] .Willie Wheeler, Joshua White "Spring in Practice"

[4] [5]. J. Sharma, Ashish Sarin "Getting started with Spring Framework

[5] Craig Walls and Manning Publications" Spring in Action (3rd Edition)

[6] [6].Craig Walls and Manning Publications" Spring in Action (3rd Edition)

[7] [7]. Clarence Ho, Rob Harrop "Pro Spring 3.

[8]

Chapters in Books:

[1] **Spring,WebMVC**, **Restful.Web Services** and **Spring Security,**Chapter 2 – Spring Frameworkbasics Chapter 3 - Configuring beans Chapter 4 - Dependency injection

. **Theses:**

[2] http://www.tutorialspoint.com/hibernate/hibernate_tutorial.pdf

Biographies and Photographs

Rizwana Unnisa

Author Name: Rizwana Unnisa

Rizwana Unnisa is currently as student of Rajarajeshwari College of Engineering in final year of Master of Computer Application. Her expertise is primarily in the domains of Wireless Sensor Networks

She obtained her BCA, from BANGALORE UNIVERCITY in 2013.

Bhanu K.N.

Author Name: Bhanu K.N.

Bhanu K N is currently working as an HOD in the Department of Master of Computer Application Her research and professional career spans for about 08 years of Teaching & research, 03 Years of Industrial Experience. Her expertise is primarily in the domains of Wireless Sensor Networks, sensor clouds, Adhoc Networks and Sensor Big data analytics.

She obtained her MCA, from IGNOU, New Delhi, M. Phil in Computer Science from VMU, Salem and presently pursuing Ph. D from Rayalaseema University, AP. She is a life member in ISTE, IAENG, SVAK and CSTA. She has published 10 papers in refereed International Journals, national and International Conferences. She has guided around 50 MCA students and 10 M.Sc. (CS) and 02 M. Phil (CS) Students for their academic projects and around 06 students have presented papers in various conferences under her guidance.

SOCIAL MEDIA ANALYTICS FOR BUSINESS ANALYSIS AND IMPROVED DECISION MAKING

Harsha Vasishta M. Sreenivasa B. R
PG student, assistant prof
Department of Computer Science & Engineering, RRCE, Bengaluru, India
harshavasishta@gmail.com, br.seenu@gmail.com

ABSTRACT : This paper proposes a system of data analytics that can be used to improve the understanding of marketing campaigns and impact of these campaigns on the decision makers in the organization. The earlier research on this topic has yielded astonishing results. However, the study was limited only to a single organization in the banking domain. Here we study the impact of social media analytics in varied fields such as real estate, charity organizations, stationery industry and beauty salons. This diversity in the nature of these organizations gives a clear understanding of the impact of social media and how this can be used to make better decisions.

Keywords - About Social Media; Digital Marketing; Analytics;

I. INTRODUCTION

Social Media has undergone a metamorphosis in the last decade. What started off as a portal of connecting groups of people on the online platform has now become a major marketing and advertising platform. People are making their buying decisions based on the brand's social media presence.

More and more companies are being driven into maintaining a noble image on the social media platforms by being prompt in replying to customer concerns and resolving issues in a quick and efficient manner. In addition to this, social media analytic data is used to streamline the marketing process and derive more results from optimized campaigns.

Currently, Facebook and Twitter are the most popular social media platforms that are used in digital marketing. New platforms such as Instagram, Vine, Snapchat and Pinterest are showing promising signs of being the next big things.

Digital Marketing started with the basic infrastructure of e-mails and web ads. With the evolution of browsers, digital marketing strategies improved a lot. Making use of cookies and device locations, it is possible to target ads that are specific to the user. Using the Moore's Law, systems can predict the buying pattern of the user.

E-commerce companies use cookies to push ads related to the product that the consumer may be interested to buy. This influences the consumer to buy the product that he finds visible most often, resulting in more sales for the companies.

With this current scenario, there is an increasing need to have a single platform to assess and utilize the information that is gathered. In this paper, we propose a web tool which uses streaming APIs to gather real-time data and then we apply clustering algorithms to make sense of patterns that are observed. This will enable the decision-makers to come to strong conclusions in a lesser time-frame. This system details the framework of the tool which encompasses many social media platforms to pull up data from every platform. Also providing a single sign-on feature, which enables seam-less connectivity between the user accounts, is proposed.

The functionality of report generation is also included to increase the ease of use. Reports are generated in portable data format. It will also include graphical representation of the data to make the presentation impressive.

II. RELATED WORK

There have been many studies in the past relating to this domain. Rozenn Dahyot, Conor Brady, Cyril Bourges and Abdullah Bulbul have proposed a system of visualizing social media analytics in their paper. Graeme Shanks and Nargiza Beckmamedova proposed a theoretical framework for Social Media analytics and business value.

These studies had a few shortcomings. The most important drawback of the first study was that it failed to encompass the multitude of platforms that are popular. Second study considers only one organization for the research. They use the framework as a lens for a case study involving a major financial corporation that used SMA as a

critical component of a major and highly successful marketing campaign.

This limits their range of use and restricts its domain. In this section, we discuss relevant literature and identify a knowledge gap in Social Media Analytics[SMA]. Social media enables users to generate content by sharing their knowledge, reviews and experiences on a plethora of issues. Social media has changed the way customers engage with brands and their services. It influences customer attitudes, perceptions and buying decisions [3]. It provides a powerful marketing platform that can be used to increase customer awareness of associated brands, products and services [1]. Also, it enables companies to improve their customer relationships through better engagement on a real time basis.

SMA involves the gathering, analysing and reporting of social media data to support efficient decision-making. SMA is being widely used in ecommerce, e-government and politics, to achieve influence and enhance customer retention, brand recognition and marketing, scale and speed, lower costs and increase flexibility.

SMA harnesses the data from social media to create a dynamic understanding of how users, organizations and brands are perceived in those social networks. SMA provides brands with a convenient platform to connect with customers and shape their perceptions through timely and targeted campaigns, responsive customer service or creating groups of mutual interest.

The Resource-Based View (RBV) has been widely used by Management Information Systems (MIS) researchers in order to understand how investments in MIS lead to organizational value-addition and competitive superiority [12]. An organization may be defined as a collection of resources that empower it to succeed and compete in the market. Resources comprise both assets and capabilities. Assets can be hardware, software, skilled personnel and data. Capabilities include competencies and practices. In order to be of strategic importance and achieve competitive superiority, resources must be valuable, rare, inimitable and non-replaceable.

Organizational resources are a critical factor of company performance. There has always been a debate over whether investment in MIS contributes to business value and whether the impact is direct. RBV theory has been used to demonstrate the effect of different types of MIS investment at different levels within organizations, including business processes, business modules and organization-wide effects [12]. Dynamic capabilities extend the RBV by 'shaping' and reconfiguring an organization's currently available resources to adapt to constantly changing technology and business environments [9]. Organizations invest in resources and learn how to use these resources over time by developing skills and accompanying practices.

The social media landscape may be explained as the intersection of the events of social media stakeholders, including specialist SMA companies and customers, and the scheme of analysis at which those events are studied [10]. We focus on social media events at the level of *management and organization*. This includes how firms manage and allocate their internal resources when using SMA to meet business targets in order to create value.

The SMA framework is based on RBV [8], dynamic capabilities [9], IT resources, organizational benefits [12], and awareness motivation [7] (see Figure 1).

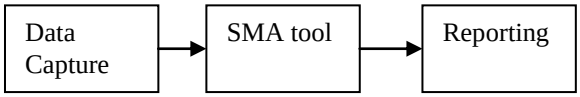


Figure 1. Social media analytics framework

The SMA framework consists of three main concepts: Data capturing, SMA tool and reporting. Data capture is the most important stage. Here the extensive data from all social media platforms are gathered to be made available for analysis. The SMA tool will cluster the data based on many parameters which are user defined. Once the data is processed, valuation results are obtained. And these results are passed to the front end to create descriptive reports.

Social media provides organizations with the ability to monitor and analyse customer views and interactions on social media platforms. In their analysis of the value of social media to organizations, Larson and Watson [7] identify three motivating consequences: information, persuasion and collaboration. While persuasion and collaboration are important for understanding the strategic use of social media, they are less relevant to the use of SMA. We focus on the analysis of social media data to generate information about customer views and interactions. The awareness is usually associated with the marketing and sales functions within organization.

III. TECHNOLOGY

Force.com is a platform as a service (PaaS) that allows developers to create add-on applications that integrate into the main Salesforce.com application. Force.com applications are hosted on Salesforce.com's infrastructure.

It is a completely cloud-based infrastructure where all the stages of SDLC are done on the cloud through a browser. Writing of code, test classes, scale testing, sandbox testing and deployment are all done on the browser.

This feature allows us to develop applications without having to install any SDK or Platforms. Another noteworthy feature is the integrating capability of the platform with APIs of different languages. In the implementation and development of the analytical tool, APIs and streaming APIs from many social media platforms were used. The huge influx of data from varied sources had to be collected and synthesized to make it useful.

IV. CASE STUDY

In this work of research, we consider the case of real-estate major Hebron Properties, Non-government organization called "Association of People with Disabilities" [APD] and Naturals beauty salons. These are picked from the list of clientele at Scion Social.

The first case of importance is that of Hebron Properties. Hebron is a popular real-estate giant and construction company catering to the elite set of customers all over southern India. Hebron decided to harness the power of SMA early in day. With a small investment they were able to create a substantial social media presence on Facebook and LinkedIn.

Due to its niche audience, targeted advertising was the most obvious choice of marketing. So Scion Social designed an important strategy. And this was to find potential customers based on the designation they hold and place of residence. A comprehensive list of CEOs, Directors, VPs and CFOs was compiled for the city of Bengaluru.

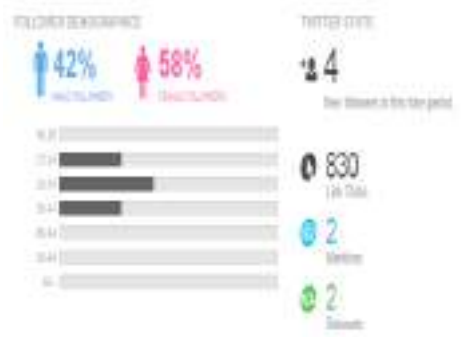
And the advertising materials were pushed only to these people over a period of time. The campaign proved to be a major success resulting in lead conversion in many occasions. The return on investment was extremely good. Scion Social's strategy of using data from LinkedIn and Facebook as a lead filtering mechanism is noteworthy.

Second case relates to that of APD. APD had a primary goal of getting donations for the organization. And a secondary goal of publicising of their activities to build credibility among funding agencies. This case is important considering the nature of domain APD is involved in.

Social media strategies were formulated in such a way that, all the activities of APD were given a run-up publicity and pictures from their campus were regularly posted to keep the audience familiar with the organisation and its working.

When a research was carried out using SMA, it was found that most of the funding to APD was coming in from USA. So a campaign was designed to attract more donations from USA. This involved ads catering to the audience comprising of average US citizens. In addition, ads were targeted at bigger NGOs working on similar lines and religious groups which were generous in contributing to social causes.

Thirdly, the Naturals chain of beauty salons entered the social media space. Using the SMA, it was concluded that the customer base of each outlet was very similar in terms of age. So the company was able to make changes in their operations to suit their young patrons. This was carried out through selfie campaigns and festive offers. In addition, the brand's connection with the customers improved a lot, resulting in customer retention. The value addition which SMA did to Naturals was immense.



V. CONCLUSION

With these observations, we can conclude that SMA plays a vital role in present day business strategies. Organizations making use of SMA are reaping rich benefits already. This study had a diverse set of cases, which gives insights of useful nature. The limitation of the study is that it does not consider new social media platforms such as Instagram, Snapchat and the likes. Multimedia data is also not considered for analysis.

ACKNOWLEDGEMENTS

We would like to thank the management and faculty of RRCE, CSE Department and also the co-founders of Scion Social, Meena Srinivasan and Siddharth Rajasekar, for their help and cooperation. The many inputs from all the members of both organizations is acknowledged with thanks. The tool used for analysis is a proprietary of Sprout Social.

REFERENCES

[1] D. Kiron, R.B. Ferguson, and P.K. Prentice, "From Value to Vision: Reimagining the Possible with Data Analytics," MIT Sloan Management Review: Spring Research Report (2013), 1-19.

[2] H. Chen, R.H.L. Chiang and V.C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact", Special Issue: Business Intelligence Research, MIS Quarterly 36:4 (2012), 1165-1188.

[3] A.M. Kaplan, and M. Haenlein, "Users of the World, Unite! The Challenges and Opportunities of Social Media," Business Horizons 53:1 (2010), p. 61.

[4] R. Kohavi, N. Rothleder, and E. Simoudis, "Emerging Trends in Business Analytics," Communications of the ACM 45:8 (2002), 45–48.

[5] N. Koudas, "Social Media Meets Business Intelligence". Retrieved 12 September, 2012, (2009) from <http://www.sysomos.com/docs/CMA2009_Sysomos.pdf>.

[6] W. Chamlertwat, P. Bhattarakosol, T. Rungkasiri, and C. Haruechaiyasak, "Discovering Consumer Insight from Twitter via Sentiment Analysis," Journal of Universal Computer Science 18:8 (2012), 973-992.

[7] K. Larson, R. and Watson, "The Value of Social Media: Towards Measuring Social Media Strategies," in Proc. of the 32nd International Conference on Information Systems (2011), China.

[8] J.B. Barney, "Firm Resources and Sustained Competitive Advantage," Journal of Management 17:1 (1991), 99-120.

[9] D.J. Teece, G. Pisano, and A. Shuen, "Dynamic Capabilities and Strategic Management," Strategic Management Journal 18:7 (1997), 509-533.

[10] P.M. Di Gangi, and M. Wasko, "Steal My Idea! Organizational Adoption of User Innovations From a User Innovation Community: A Case Study of Dell IdeaStorm," Decision Support Systems 48:1 (2009), 303-312.

[11] Kurniawati, G. Shanks and N. Bekmamedova, "The Business Impact of Social Media Analytics," Proc. European Conference on Information Systems 2013, Paper 52.

[12] S. Aral, and P. Weill, "IT Assets, Organizational Capabilities and Firm Performance: How Resource Allocations and Organizational Differences Explain Performance Variation," Organization Science 18:5 (2007)

A FRAMEWORK ON PENETENTIARY CONSERVANCY

Syeda asra samreen, Bhanu.K.N
Department of MCA, Rajarajeshwari College of engineering, Bangalore, India
syedaasrasamreen@gmail.com, bhanu.kn@gmail.com

ABSTRACT-This framework is developed for proper organization and maintains the activities related to prisons. The purpose is to design a system that provides basic functionalities for storing and accessing the information related to prisons. This system allows one to access and search for the required information very easily. The system allows generating various reports of criminals. This system is the reinstate of the manual system. Incarceration in prison, also known as a penitentiary, is the sentence that courts most frequently enforce for vital crimes, such as criminal act. For lesser offense, courts mostly inflict short-term internment in a prison, or correction center as mentioned in [7].

Keywords – Conservancy, Demographic, Nominal roll, Parole, Penitentiary.

I. INTRODUCTION

T he purpose of this paper is to develop an application for maintenance of the prison system. This application consists of various registers, reports for effective operations of prisons. And police/employee can see crime/criminals reports for their purpose. A prison is a place in which people are kept forcefully in a prison as punishment for a crime they. Prisons are buildings or institutions, in which criminals get punished for the crime they have committed. Prisons are also called as a remand center for the criminals. The punishment for the criminals is based on the crime they have committed and prisons are also called as retention center as mentioned in [6].

In existing system it is tedious process because all the work was done manually. Less secured so the proposed system is introduced to overcome the problems of existing system. In a proposed system the work is automated which is easy and secure to access. This system is fully automated and designed in a simple so that it can be accessed easily. Authentication is given so that only the registered members can access this application. Without the username and password the employees of jail cannot access this application.

II. SYSTEM DESIGN

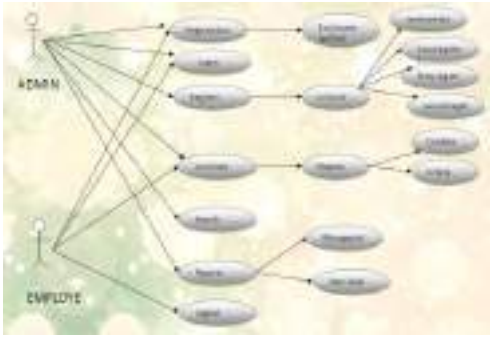


Fig2. Use case diagram

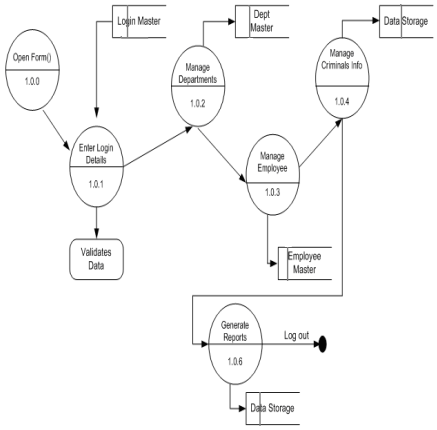


Fig3. Level 1 Data flow diagram for admin

Fig2. And Fig3. Discuss about the role of admin and employee in this application

In this paper the proposed system is divided into two parts

1) Administration

2) Employee/police officer.

Prisons are also called as corrections center for the criminals as mentioned in [7]. In Fig3.Admin is the super user in this system .Admin can register an employee based on their designation and department. Admin can view all the information of users and criminals. Admin can view all the registers of criminals like parole register, case register, duty register etc...And admin can make changes to these registers. He can easily know about the criminal which reason he has arrested or other related information.

Admin can view the Criminal case details. So that he can know which case is going on the particular criminal Admin can view the Criminal court details so that he can know in which court the particular criminal's case is going on. Through court details, He can know when the next hearing of particular criminal. Admin can also view/modify the employee details and can assign duties for the employs. He can view registered employee details where he can know when employee has join and other necessary information of employees. Admin can add department details like, Jail, Account, and Mess etc. He can add different type of designations also.

From Fig2. Employee should login first by giving valid username and password to access the application. He can view his profile and can modify necessary changes .Employee can register a new criminal and maintains all the information of criminals. He can assign a task or work to a criminal. He can update/delete the criminal's information. Employee should maintain the records of all the criminals. The details include like case details, duty details, and released criminals details.

Salary details of an employee are stored in salary database. Employee can only view his salary details, per month salary, total amount calculated after increment and decrement per month. He can just view his total salary.

III. GRAPHICAL USER INTERFACE

Fig4. Admin registers new employee:

Fig4. Explains how the admin adds new user. Admin registers new employee by giving all the details of an employee and assigning a password for employee. This form will be filled by the admin because admin has the power to assign id and password for the employees. So if the admin needs to add a user he should register first and assign id and password to the employee so that he can access the application. The details include the name of employee, his personal details like date of birth, marital status etc.

Fig5. Admin views employee duties

Fig5. Discussed how the admin views the details of employee. The employee duty details assigned by admin can be viewed by admin and the employee. But only admin can make changes for the duty details. These are the details of employee duties assigned by the admin. Each employee will have a respective duty. The duty date and time is allotted by the admin. Employee cannot change the duty details only admin have a power to modify, he can change the duties. Duty details include duty id, duty in charge and remarks.

Fig6. Adding new criminal

Fig6. Discusses about how the employee adds a new criminal. Adding new criminal can be done by an employee. He can add all the details of a criminal. All the details will be stored in database. Employee adds all the details of criminals like what crime he has made, who is the lawyer of that criminal, when is the hearing and all the court details and so on. These details will be stored in the backend. The details to be entered in the front end that is graphical user interface. Employee can modify the details of criminal if the criminal has already released so the changes to be done in database. Each criminal will be assigned a unique id so that he can easily be identified.

Fig7. Criminal released details

Fig7. Explains the how employee adds the released details of criminal. Employee should maintain the details of the criminals who are already released or the criminals to be released soon, their date of releasing and reason for released and remarks if any. These details should be maintained by the employee. The released criminals are assigned release id through which the employees can identify who is the released. The reasons for releasing all should be maintained by a particular employee who is assigned for this duty. Employee should maintain the details of the criminals who are in jail and to be released soon and later.

Fig8. Criminal duty details

From Fig8. We discuss how the criminal duty details will be added. In this employee will also maintain the details of criminal duties. Employee should assign duties for the criminals by assigning duty id, day of duty, time of duty and place of duty of a criminal. Jail is a place where the criminals should work over there and they will be paid for it. These details will be assigned by the employee when and where a criminal should complete his duty and time of their duty.

IV. CONCLUSION

This framework helps in implementation of central and state prisons. Retaining crucial information in the files and manuals is full of risk and a tedious process, so to overcome the problems of the manual system the new system switched the manual system to a software solution. It automates the operations carried out in the prisons, and maintenance of prisons in a proper and automated way.

VI. ACKNOWLEDGMENT

The authors of this paper are thankful to the Management, RRGI, Bangalore and Principal, RRCE, Bangalore for their continuous encouragement and support throughout the work

REFERENCES

[1] Robert Schildt, "Java the Complete Reference
[2] Core Servlets & Java Server Pages, 2nd Edition Vol 2
[3] MySQL, Paul DuBois, 5th Edition
[4] High Performance MySQL, Baron Schwartz, Peter Zaitsev, Vadim Tkachenko , 3rd Edition
[5] Mohr Yasin, N & Syed A. Kadir, B. S. L. (2008). Web advertising beliefs and attitude: Internet users'view. The Business Review.
[6] https://en.wikipedia.org/wiki/Prison
[7] http://www.ala.org/offices/sites/ala.org.offices/files/content/olos/prison_facts.pdf

Syeda Asra Samreen is currently as PG student of Rajarajeshwari College of Engineering pursuing final year of Master of Computer Application.

Internet of Things for Environmental Monitoring

,Bhoomika.K.N[#2], Deepa.C[#3], Rashmi.R.K[#4], Srinivasa.R[*1]
[2,3,4]UGstudents, [1]Associate Professor, Dept of Studies in CSE,RRCE, Bangalore

ABSTRACT- Increasing population in cities demands satisfactory provision of services and infrastructure for the city's residents and visitors. The deployment of information and communication technologies to accomplish this objective presents an opportunity for the progress of cities, where management of city and citizens are given access to real-time information about the environment. This paper presents the structure for environmental monitoring using the Internet of Things (IoT). The structure comprises the complete information system from the sensor level to data management and cloud-based information about the environment. This Iot vision is applied to temperate, waste management and vehicle parking to demonstrate a method for existing systems that can be adopted for the enhancement and delivery of services.

Keywords—Internet of Things (IoT), Temperature, waste management, vehicle parking, MQTT.

I. INTRODUCTION

The Internet of things (IoT) is network of objects, in which the objects of everyday life are embedded with microcontroller, sensors and software that enables these objects to collect and communicate data with one another and the users, becoming the essential part of the internet. The IoT model, aims at making the Internet even more persistent. Furthermore, by enabling easy access and communication with a wide range of devices such as, for example, home appliances, surveillance cameras, monitoring sensors and so on, the IoT is implemented for the development of applications that makes use of the enormous amount and the data generated by such objects provide services. This method finds application in many different areas, such as home automate, mobile healthcare, traffic management and many others [1].

In this scenario, the application of the IoT paradigm to a city environment is of particular interest, as it responds to the governments to adopt information and communications technologies (ICT) solutions in the management of public affairs, thus it is called Smart City concept [2]. Though there is no yet formal and widely accepted definition of "Smart City",the aim is to make a better use of the municipal resources, increasing the quality of the services presented to the residents, while decreasing the operational costs of the municipal administrations. This objective can be tailed by the deployment of an IoT, i.e., a communication infrastructure that provides cohesive, simple, and inexpensive access to a overabundance of public services, thus unleashing potential interactions and increasing clearness to the citizens. An IoT has numerous benefits in managing and optimizing traditional services, such as transport and parking, lighting, observation and maintenance of public areas, protection of cultural heritage, garbage collection, hospitals, and school. Furthermore, the accessibility of different types of data, which is collected by a persistent IoT, may also be used to take advantage to increase the clearness and promote the actions of the local government toward the residents, improve the awareness of people about the status of their town, stimulate the active participation of the residents in the management of public administration, and also stimulate the building of new services provided by the IoT [3]. Therefore, the presentation of the IoT standard to the City is particularly to regional and regional administrations that may become the early implementation of such technologies, thus acting as catalyzes for the implementation of the IoT paradigm on a wider scale.

The paper is organized as follows: Proposed system, especially from the perspective of environment is first given in Section II. We then present the details of the IoT implementation for environmental monitoring in Section III, the results for the proposed system in Section IV. We also present the summary and future thoughts in section V and section VI respectively.

II. PROPOSED SYSTEM

IoT has various of application in this paper we are discussing about the temperature, vehicle parking and waste management. These are the few issues which we are facing regularly in our day today life with the use of IoT we can provide solutions to these problems.

A. Temperature

The IoT plays a great role in the development of future smart cities. For instance, Air Quality Eggs can be found across America, Western Europe and East Asia, and may finally play a role in developing countries with the most rapid urban population growth and peak rates of pollution. This is a community-led air quality sensing network that allows anyone to collect very high contents of Nitrogen-di-oxide (NO2) and Carbon monoxide (CO) concentrations of their surroundings. These two gases are the most suggestive elements related to urban air pollution that are sense-able by reasonably priced, DIY sensors. The temperature plays a very crucial role. The temperature varies from time to time. The temperature in the day times is completely different in the night. In paper gives an idea of how a temperature can be identified for the city using IoT concept.

B. Waste Management

Waste management is a main issue in many growing cities, due to both the cost of the service and the problem of the storage of garbage accumulation. A deeper penetration of information and communications technologies solutions in this field may result in the savings and inexpensive and environmental advantages. For instance, the use of intelligent waste containers, which identify the level of load and allow for an optimization of the collector trucks route, can reduce the cost of waste collection and improve the quality of recycling [4]. To realize such a smart waste management service, the IoT will connect the devices, i.e., intelligent waste containers, to a control centre where optimization software process the data and determines the optimal management of the collector truck.

C. Vehicle Tracking

The vehicle tracking facility is based on road sensors and intelligent displays that direct drivers along the ZANELLA et al.: INTERNET OF THINGS FOR SMART CITIES 25 best path for parking in the city [5]. The benefits deriving from this service are various: faster time to locate a parking slot means fewer CO emission from the car, lesser traffic congestion, and happier citizens. The vehicle parking facility can be directly integrated in the IoT infrastructure. Furthermore, by using communication technologies, such as Radio Frequency Identifiers (RFID) or Near Field Communication (NFC), it is possible to understand an electronic confirmation system of parking permits in slots reserved for residents or disabled, thus offering a better service to residents that can legitimately use those slots and an efficient tool to quickly spot violations.

III. IMPLEMENTATION

IoT service provides a simple but powerful capability to interconnect different kinds of devices and applications all over the world. IoT service acts as MQTT broker and is thus responsible for distributing message to connected clients. Devices and applications communicate with MQTT broker using MQTT protocol.

A. Temperature Architecture

This paper discuss about the temperature variations in and around the place using the data received from the satellite.
There are three entities defined in our system that is cloud devices and app as shown in the Fig.1.

Cloud: Cloud becomes prevalent, an increasing amount of data is been stored in the cloud and shared by the devices with specific privileges, which defines the access rites of the stored data. The user gets an API Key, the password and the device purpose for which it is registered. Once he is registered to the cloud. Using these API Key and the password he could access the data received from the devices.

Devices: Devices can act as sensors for the purpose of receiving or to transmit the data to the cloud. The devices can be anything which can sense the data.

Apps: Applications are programs that consume the information received from those devices.
MQTT: Message Queue Telemetry Transport/Things transport. In IoT things are nothing but the devices like Buildings, systems etc. Telemetry is an automatic device which is used to measure and transmit the data. It can be considered as the secrete behind the IoT service. MQTT is a simple lightweight, publish/subscriber messaging protocol on top of TCP/IP protocol

JSON: JavaScript Object Notation. It acts as a bridge between the physical objects and the web sensors.

B. Waste management and Vehicle Tracking

This paper also discuss about the waste management and vehicle tracking using the data received from the satellite. There are three entities defined in our system that is cloud devices and app as shown in the Fig.2.

Cloud: Amount of data is been stored in the cloud and shared by the devices with specific privileges, which defines the access rites of the stored data. The user gets an API Key, the password and the device purpose for which it is registered.

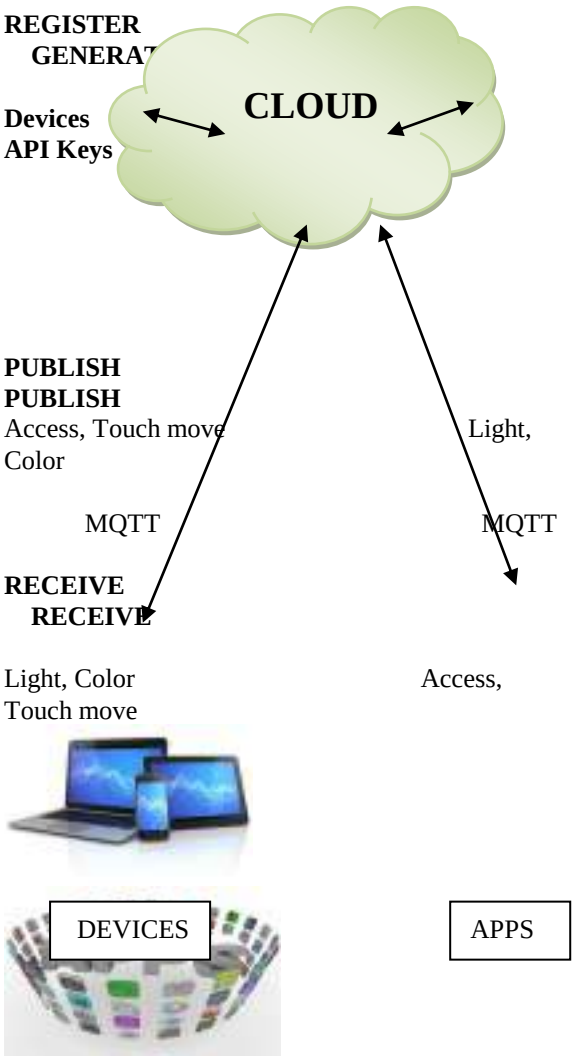


Fig. 1 Architecture for Temperature IoT

Once he is registered to the cloud. Using these API Key and the password he could access the data received from the devices.

Devices: Since now a day most of the devices are GPS enabled, it is very easy to track the exact locations of the vehicles. Since in our day to day life, we do not know the exact time when the waste truck arrive to collect the garbage from every particular area. The details of the waste truck and the vehicle to be tracked are based on the latitude and longitude of the particular region.

Apps: Programs that receives data from the devices.

As in the below description the app is used to access the data from the cloud.

Node red is a simple open source visualization tool that connects devices for the IoT. Node red has been developed in node.js, a server JavaScript widely used in IoT pits and can be run in cloud.

Notice that the data collected from these devices are accessed by the satellite which in turn sense these accessed data to the base station. The data sent to the base station is sent in the form of raw data. The cloud filters the raw data and the filtered data can be accessed by the app. Both apps and devices place a very important role in publishing the collected data to receive the data.

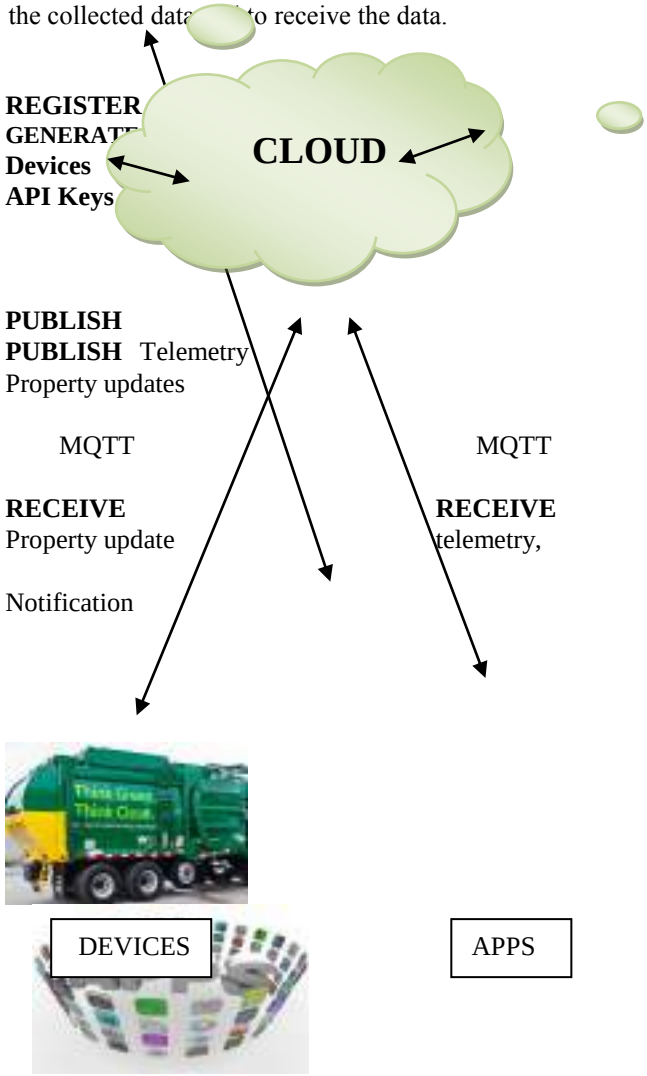


Fig. 2 Architecture for Waste management and vehicle Tracking IoT

IV. RESULT

Basically we have to get connected with the cloud to retrieve the data from the cloud. We can see the login page in Fig. 3, where the user has to login with the particular API Key and the token which is given by the cloud so that privacy is maintained to access the server.

In Fig. 4 we can see the temperature variation. We'll be getting the second to second update of the temperature. And through the graphs and meter we can get to know the variation in temperature. If the

temperature is less than 30, the indication will be in green color. If the temperature lies between 31 and 75, we can see that the color changes to yellow and if the temperature is above 75 then the color changes to red.

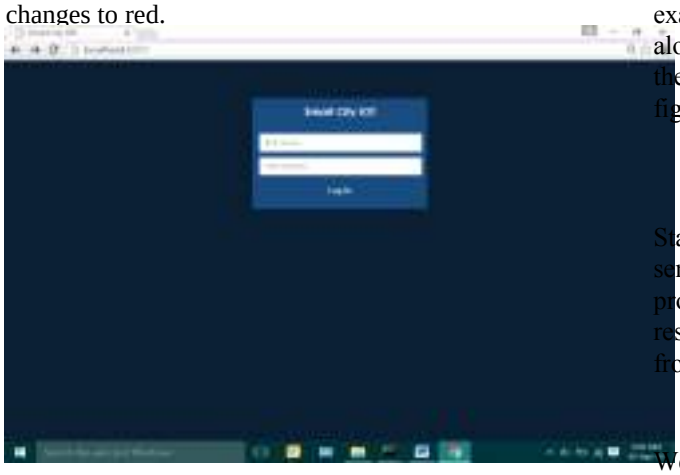


Fig. 3 Login Page

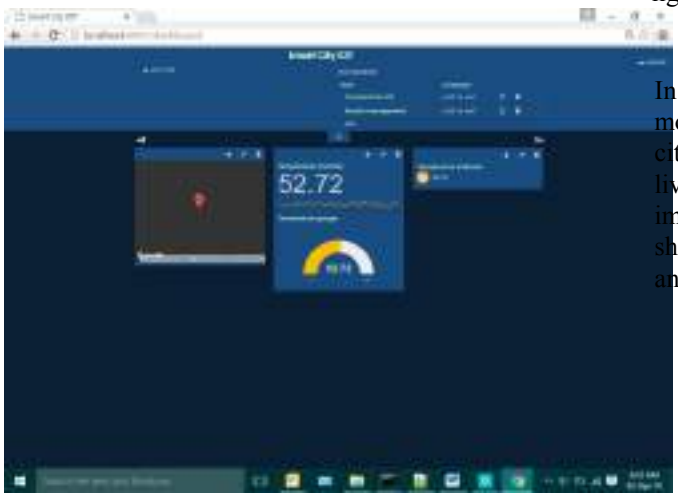


Fig. 4Result of Temperature.

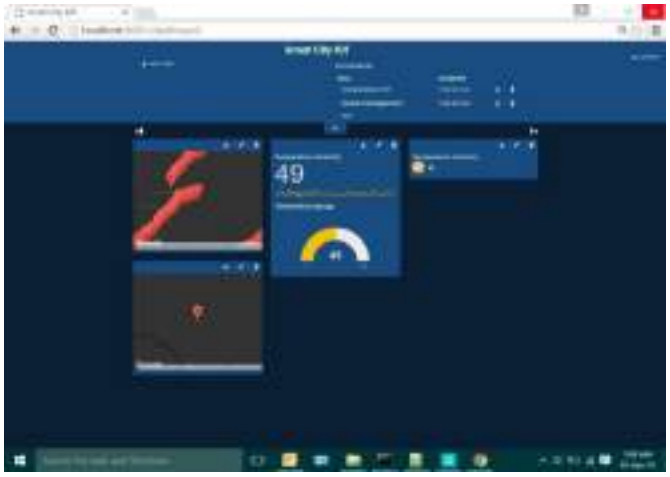


Fig. 5 Overall result of the project

In fig. 5 we can see that we are linking to Google map for tracking the cab of the waste collector. And we can also see that the garbage collector truck has been tracked so that the user will get to know the exact location of the garbage collector truck. Also along with the garbage collector truck we can see the temperature as well as vehicle tracking in this figure.

V. SUMMARY

Start the IoT starter app in device, it starts to collect sensor data. From the device it goes over MQTT protocol to IoT cloud and there you can access via rest API. The data collected are accelerometer data from device.

VI. FUTURE WORKS

We can also implement this in android watches, attached to other IoT concepts like home security, structural health, city energy consumption, smart lighting and so on.

VII. CONCLUSION

In this paper, the notation of environmental monitoring using IoT was proposed to monitor the city from the damages cost which may affect the living of the citizens. As a proof of concept we implemented a IoT of a proposed system. We showed that our system can access the data without any embedded system.

REFERENCES

[1]P. Bellavista, G. Cardone, A. Corradi, and L. Foschini, "Convergence of MANET and WSN in IoT urban scenarios," IEEE Sens. J., vol. 13, no. 10, pp. 3558–3567, Oct. 2013.

[2]H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, "Smart cities and the future internet: Towards cooperation frameworks for open innovation," The Future Internet, Lect. Notes Comput. Sci., vol. 6656, pp. 431–446, 2011.

[3]D. Cuff, M. Hansen, and J. Kang, "Urban sensing: Out of the woods," Commun. ACM, vol. 51, no. 3, pp. 24-33, Mar. 2008.

[4]T. Nuortio, J. Kytöjoki, H. Niska, and O. Bräysy, "Improved route planning and scheduling of waste collection and transport," Expert Syst. Appl., vol. 30, no. 2, pp. 223–232, Feb. 2006.

[5]W. Kastner, G. Neugschwandtner, S. Soucek, and H. M. Newmann, "Communication systems for building automation and control," in Proc. IEEE, Jun. 2005, vol. 93, no. 6, pp. 1178–1203.

[6] R. T Fielding, "Architectural styles and the design of network-based software architectures," (The Representational State Transfer (REST)) Ph.D. dissertation, pp 76-85, Dept. Inf. Comput. Sci. Univ. California, Irvine, 2000. [Online]. Available: http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm.

[7] A. P. Castellani, N. Bui, P. Casari, M. Rossi, Z. Shelby, and M. Zorzi, "Architecture and protocols for the Internet of Things: A case study," in Proc. 8th IEEE Int. Conf. Pervasive Comput. Commun. Workshops

[8] (PERCOM Workshops), 2010, pp. 678–683.

[9] A. P. Castellani, M. Dissegna, N. Bui, and M. Zorzi, "WebIoT: A web application framework for the internet of things," in Proc. IEEE Wireless Commun. Netw. Conf. Workshops, Paris, France, 2012.

[10] Z. Shelby, K. Hartke, C. Bormann, and B. Frank, Constrained applica-tion protocol (CoAP), draft-ietf-core-coap-18 (work in progress), s.l.: IETF 2013. [Online]. Available: http://tools.ietf.org/html/draft-ietf-core-coap-18.

[11] A. Castellani, S. Loreto, A. Rahman, T. Fossati, and E. Dijk, Best practices for HTTP-CoAP mapping implementation, draft-castellani-core-http-mapping-07 (work in progress), s.l.: IETF 2013. [Online]. Available: https://tools.ietf.org/html/draft-castellani-core-http-mapping-02.

[12] S. Deering and R. Hinden, Internet Protocol, Version 6 (IPv6) Specification, RFC2460, s.l.: IETF Dec. 1998. [Online]. Available: https://www.ietf.org/rfc/rfc2460.txt.

[13] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, Transmission of IPv6 packets over IEEE 802.15.4 networks, RFC4944, s.l.: IETF Sep. 2007. [Online].Available: http://tools.ietf.org/html/rfc4944.

[14] J. Hui and P. Thubert, Compression format for IPv6 datagrams over IEEE 802.15.4-Based Networks, RFC6282, s.l.: IETF Sep. 2011. [Online]. Available: http://tools.ietf.org/html/rfc6282.

[15] IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs), IEEE Standard 802.15.4-2011.

[16] IEEE 802.15 WPAN Task Group 4e (TG4e), IEEE Standard 802.15.4b, 2014.

[17] A. P. Castellani, M. Gheda, N. Bui, M. Rossi, and M. Zorzi, "Web services for the Internet of Things through CoAP and EXI," in Proc. IEEE Int. Conf. Commun. (ICC 2001), Kyoto, Japan, 2011.

[18] P. Casari et al., "The Wireless Sensor networks for city-Wide Ambient Intelligence (WISE-WAI) project," MDPI J. Sensors, vol. 9, no. 6, pp. 4056–4082, Jun. 2009. [Online]. Available: http://www.mdpi.com/1424-8220/9/6/4056.

An Generic Cloud Framework for Cloud Based Applications

Arvind Kumar M, Mr T.Auntin Jose
Computer Science & Engineering, RRCE, Bengaluru, India
kmarvind@gmail.com, Auntin123@yahoo.com

***Abstract*—**This paper focuses on solving the problem of cloud application developers facing while using the available Clouds in their cloud application development process. There are multiple cloud infrastructures available in the market. The Cloud Application developers can't be familiar with all the clouds available in the market. Our paper addresses this issue, and talks to facilitate the cloud application developer to provide an generic abstract layer which hides the back end cloud infrastructure and with a uniform API layer to make use in their cloud applications. Generic Cloud Frame work will provide the facility to common developers to have a minimum knowledge about the cloud and solve the various problems. With the help of Generic Cloud Framework, the developers can provision and develop various applications like provisioning for DB management systems, Medical Systems, Educational systems, Agriculture systems, using an "Infrastructure as a Service" (IaaS) cloud. To solve this problem in a generic way by providing the domain specific API Libraries with uniform interface to back end cloud infrastructure.

Keywords— *Cloud, IaaS; PaaS; Library; Analytics;*

I. INTRODUCTION

Cloud computing is not an technology now a days. But it is becoming and a valuable and important which is becoming an fundamentally changing the way we use and develop on-demand applications. As you we know, Linux & open source provides the foundation for the cloud (either for public or private infrastructures). We explored the cloud anatomy, its architecture, and the open source technologies used to build these platforms. Cloud computing is commonly considered as "the greatest sliceable thing in IT infrastructure ", as it allows the computing power can be utilized to behave similar to a generic utility that is always available to user on their need basis from anywhere. This utility-orientation about the hardware and software usage makes the cloud computing as an potential transformative technology. The model of pay-as-you-go service makes it an essential economic savvy by reducing the capital expenses and reducing the cost of ownership over time.

Most of the applications are good candidates for adapting to the cloud — if the applications are difficult to configure and manage, then they are the definite candidates for hosting them in the cloud. Many applications have been already ported on virtualized hardware provided by IaaS, SaaS or PaaS providers. The key factor need to be keep in mind is, cost-effectiveness in deployment of cloud applications remain open. One such key factor involves application optimization, which basically assumes the availability of Hardware, Software resources, configuration of them which will result in performance optimization. This ignores the OPEX (operational cost) of running the applications. The pay-as-you-go model used in cloud computing, however, attracts the large number of users to consider hosting the applications on the Cloud, to reduce operational costs and consider increasing the performance.

In this paper, we discuss how any kind of users can easily implement their cloud applications by minimizing the detail knowledge of the backend cloud. This reduces the operational & development cost of a Cloud Applications. That is to say, the Libraries provided by the Generic Cloud Framework will be tapping all the requests from the Cloud applications and translating to back end cloud specific calls to use the specific resource types from physical available cloud at their disposal. We propose a resource provisioning framework with common generic API libraries which identifies the heterogeneous cloud environment, resources, virtual machines etc. that can collectively abstracts the user from Cloud specific knowledge. The requirement for the developer is to have detailed knowledge about the APIs provided by the library. At run time these calls will be translated to make best use of resources by intelligently routing the incoming queries to specific cloud environment.

The work is still in progress. In this work in progress paper, we describe one domain and few set of API libraries to use limited cloud resource for provisioning and utilizing them. We can be termed it as a black box provisioning as the user no need to have any insight or the resource usage of the back-end cloud. An sample application on desktop and handheld device is developed using the Generic Cloud Frame work APIs to demonstrate the end to end flow of the data from domain user to cloud.

We formulated these solutions to help the cloud application developer for packaging and tackling their applications with greater ease and efficiency.

II. SYSTEM STRUCTURE

Cloud Infrastructure Services. Our model system assumes an cloud infrastructure which is similar to the offerings made from various IaaS providers, Cloud vendors provide access to the pre-configured computing power they could be virtual machines or cloud servers on which users can remotely install and run their software. Available server types are categorized by the resources they provide to the end user, such as CPU, I/O Band Width, and Physical disk space and system memory size. Cloud resources are usually offered through a "pay-as-go" price model, they can be rentable for a certain predefined period, probably for an hour, and this period cost is fixed. The cost depends only on the server configuration and not on the server utilization over the time. IaaS clouds are also will provide the services for data storage and transfer. Usually the additional request charges will cost the user, For example, Amazon's EC2, S3, & EBS services offers persistent storage space and charged based on the per request basis for each I/O operation. This work considers the charges only for the Servers and I/O operations. Regardless of pricing model used, it is the responsibility of the IaaS user to provision appropriately to use its resources. Based on our estimation the workload expected to be by hosting the application.

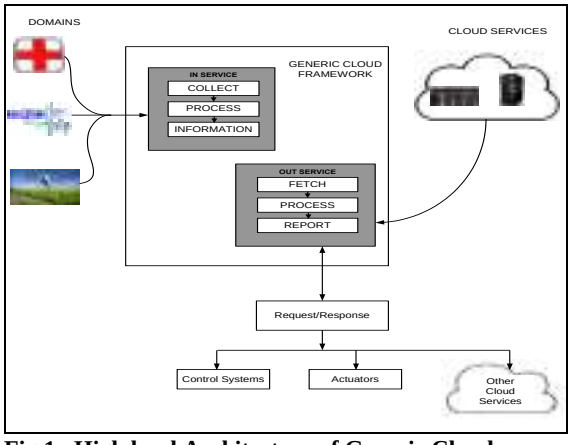


Fig.1. High level Architecture of Generic Cloud Framework

Domain Specific Components. Cloud-applications for specific domain need specific services. Based on the Hardware, Software required for that domain need to have customized APIs to configure Hardware, APIs for domain specific operations, APIs for Domain specific analytics etc. We pre-assume that the application writer is a domain expert rather a cloud expert. Each domain will have a specific library and the application developer need to understand the library instead of the cloud infrastructure. The APIs reside on API server with wrapper routines to the back-end cloud infrastructure. The (Figure 1) shows over a high level architecture of the system. The application development for specific domain needs appropriate library.

In-Service Module. This module will have the interface to cater the Domain specific community and Cloud service. It will receive the Client request and translate it into the back-end cloud specific call. The calls received from the client are un-wrapped; the parameters are analysed and forwarded to specific cloud utilities and infrastructure modules provided by the cloud infrastructure provider calls. For meaningful, the in-service module receives the request, process the request and at the end, it will send it to the cloud for servicing the request. Based on the service requested the expected distribution of requests across the different service handlers also happens based on this module. This module provides facilities for methodical representation and facilitates to generate reports for various purposes.

Out-Service Module. Provisioning the data collected in the cloud from In-Service module to end users on need basis for reference, for predictions, etc. The Out-Service module is responsible for representing the data to the users. Again to fetch the data from cloud, the user will have domain Specific APIs . The client uses these set of APIs and read the data for representation, dash boarding, Analytics, decision making purposes. The out-service module provides the APIs to fetch the data either on the desktops or on any hand held devices.

The APIs interactions could be represented as shown below picture.

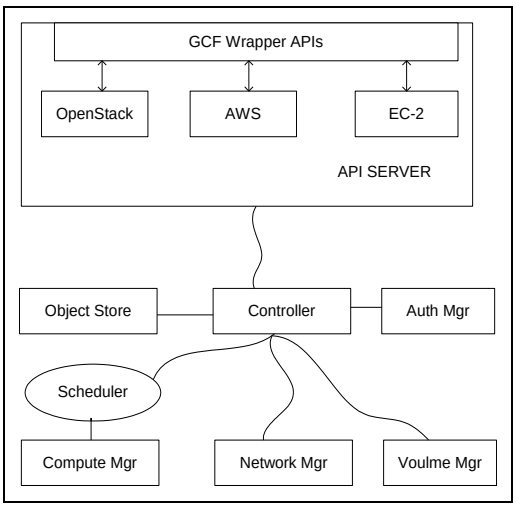


Fig. 2. The Interfaces interaction

III. EXPERIMENT CONFIGURATIONS

Throughout this paper we reference the results using the proof-of-concept for our framework deployed for an application in an agriculture environment, which involves Hardware modules, Software modules, Google services, Amazon Web Services [1]. We deployed the Hardware to read the vegetative index of a particular geographical area's crop and upload the data using In-Services of Generic cloud framework. In-turn this data could help to analyse the health of a particular crop in a specific geo-locations. It can be used
to advice the agriculture community for predictive analysis, like what could be the proactive health care of the crops by using proper ratio of Urea, Nitrogen fixes and appropriate pesticides at right time and right composition. The hardware module used for reading the crop health, and is uploaded to the cloud along with the crop's vegetative index, and geo co-ordinates. In this Proof Of Concept the cloud environment used is Amazon Web Services. The HW used are, PIC Microcontroller, Vegetative index reading sensor, GSM and GPRS 908 module 908. All these components are interconnected as below,

Geolocation based Vegetative Index					
Longitude	***Latitude***	***Humidity***	***Temperature***	***Vegetative Index***	***City / Feedback***
12.91N	77.92 E	28	34	5F	B'lore
13.08N	80.27 E	70	30	3B	Chennai
17.38N	78.41 E	23	36	2C	Hyderabad
19.07N	72.87 E	62	31	1A	Mumbai
22.51N	88.36 E	66	31	32	Kolkota

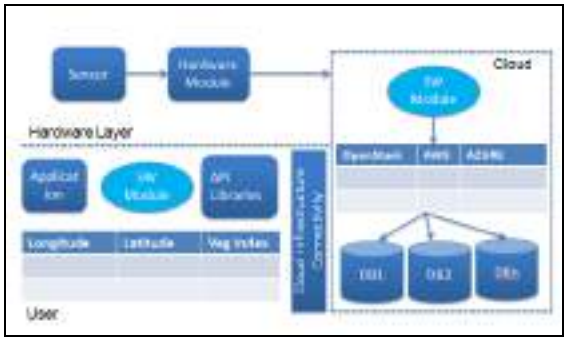


Fig. 3. Generic Cloud Frame work, POC configuration

This POC generates the reports in the form of crop's geo location, it is health indicator and environment parameters are also can be captured, if the necessary sensors are used to collect the data from the field. We also can extend this and optimize to produce estimates for any specific crops, analytics modules will give the usage of each micro components and ratio analysis for demographic related parameters can be defined. We correlated this with actual resources consumed without using these modules in real field. The consistency and high availability concepts can also be supported and achieved.

IV. GENERIC CLOUD FRAMEWORK REPORTS

The report generated by this proof of concept is as mentioned

Table .1. Sample Report structure

In the above given table. The sensor information is captured and could be used for the analytics purpose. This data set information will be will be uploaded into the cloud database. The user can fetch, process and analyse with the help of APIs provided by the GCF.

V. CONCLUSION

The outcome of the generic cloud framework is in multi-fold. It directly helps the user who would like to use the cloud environment for their specific application development activities. Thus Generic Cloud Framework gives users the speed; simplicity and control they need to develop and deploy applications easily and faster. It provides different specific API layers and infrastructure customized to their development environments.

Over all benefits of this framework are listed as below,

- Enable users to have real time data.
- Application development will be very quick since developer need not be aware about cloud services.
- Reuse of the frame work for multiple purposes.
- Data analytics can be done by understanding one single library on various domain data.
- Easy to maintain the application code.

Enable users to have real time data. User can gain flexibility & control over their choice of languages, frameworks, application services and clouds. Portability is ensured, and vendor lock-in can be eliminated. As the global industry standard for PaaS is open source technology, we operate under the open governance by contribution model. The customer who generates the data can be processed and stored in a defined common standard format, so that it can be used at later point of time by the future customer for their analysis purpose.

Quick Application development. By connecting human brainpower and computing power, we reduce development times and accelerate how applications are designed, developed and delivered. Empowering the companies to cut weeks and months durations once needed to develop and ship new applications to just days, hours, even minutes, Generic Cloud Framework makes faster easy.

Reuse of the frame work for multiple purposes. Built on hardened production infrastructure for global enterprises, the GCF platform ensures scalable micro services and continuous deployment, faster cycle time and higher reliability. Everything is fully scalable, including the platform itself. The framework developed can be reused for various purposes.

Data analytics can be done on various domain data. Built on hardened production infrastructure for global enterprises, the platform ensures scalable micro services and continuous deployment, faster cycle time and higher reliability. Everything is fully scalable, including the platform itself. The data generated can be tiered and utilized for future analytics purpose. Only data also can be rendered to the customers who are looking for the for R&D purposes in multiple domains.

Easy to maintain the application code: Built on hardened production infrastructure for global enterprises, the platform ensures scalable micro services and continuous deployment, faster cycle time and higher reliability. Everything is fully scalable, including the platform itself.

ACKNOWLEDGMENT

This work has been partially incorporated by the open stack group as a part open source and getting customised for commercial use under various brand names like Open Shift.

REFERENCES

[9] Amazon Web Services, http://aws.amazon.com/.

[10] D. J. Abadi. Data management in the cloud: Limitations and opportunities. IEEE Bulletin on Data Engineering, 32(1), 2009.

[11] A. Aboulnaga, K. Salem, A. A. Soror, and U. F. Minhas. Deploying database appliance in the cloud. IEEE Bulletin on Data Engineering, 32(1), 2009.

[12] M. Ahmad, A. Aboulnaga, S. Babu, and K. Munagala. Modeling and exploiting query interactions in database systems. In CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, pages 183–192, New York, NY, USA, 2008. ACM.

[13] M. Brantner, D. Florescu, D. Graf, D. Kossmann, and T. Kraska. Building a database on s3. In SIGMOD, 2008.

[14] E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good. The cost of doing science on the cloud: the montage example. In SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing, pages 1–12, Piscataway, NJ, USA, 2008. IEEE Press.

[15] R. P. Doyle, J. S. Chase, O. M. Asad, W. Jin, and A. M. Vahdat. Modelbased resource provisioning in a web service utility. In USITS, 2003.

[16] A Highly Scalable Transactional Multi-Tier Platform as a Service from Director Distributed Systems Lab (LSD), Univ. Politecnica de Madrid (UPM) Grant Agreement number: 257993 – 2013

[17] Efficient & scalable Para virtual I/O system – 2013 USENIX Annual Technical conference

[18] Sukhpal Singh and Inderveer Chana, "QoS-aware Autonomic Resource Management in Cloud Computing: A Systematic Review","ACM Computing Surveys" , Volume 48, Issue 6, pp. 1-39, 2015

[19] Rajkumar Buyya, Rodrigo N. Calheiros, and Xiaorong Li. "Autonomic cloud computing: Open challenges and architectural elements", *In* Proceeding of the Third International Conference on Emerging Applications of Information Technology (EAIT). (2012), 3-10, IEEE.

[20] Kaloxylos, Alexandros, Robert Eigenmann, Frederick Teye, Zoi Politopoulou, Sjaak Wolfert, Claudia Shrank, Markus Dillinger, "Farm management systems and the Future Internet era", *Computers and Electronics in Agriculture*, 89 (2012): 130-144.

[21] Ranya Elsheikh, Abdul Rashid B. Mohamed Shariff, Fazel Amiri, Noordin B. Ahmad, Siva Kumar Balasundram, and Mohd Amin Mohd Soom, "Agriculture Land Suitability Evaluator (ALSE): A decision and planning support tool for tropical and subtropical crops", *Computers and Electronics in Agriculture*, 93 (2013): 98-110.

[22] Raimo Nikkilä, Ilkka Seilonen, and Kari Koskinen, "Software architecture for farm management information systems in precision agriculture", *Computers and Electronics in Agriculture*, 70(2) (2010): 328-336.

[23] C. G.Sorensen, S. Fountas, E. Nash, Liisa Pesonen, Dionysis Bochtis, Søren Marcus Pedersen, B. Basso, and S. B. Blackmore, "Conceptual model of a future farm management information system", *Computers and Electronics in Agriculture*, 72(1) (2010): 37-47.

[24] Zhao Ruixue, "Study on Web-based Agricultural Information System Development Method", *In Proceedings of the Third Asian Conference for Information Technology in Agriculture, China*, 601-605. 2002.

[25] C. G. Sørensen, Liisa Pesonen, D. D. Bochtis, S. G. Vougioukas, and Pasi Suomi, "*Functional requirements for a future farm management information system*", Computers and Electronics in Agriculture, 76(2), (2011): 266-276.

[26] Yuegao Hu, Zhi Quan, and Yiyu Yao, "Web-based Agricultural Support Systems", *In Proceeding of the Workshop on Web-based Support Systems*, 75-80. 2004.

[27] The Open Compute project—energy efficiency. (Available from: http://opencompute.org/about/energy-efficiency/) [Accessed on 21 November 2012].

[28] Hermenier F, Lorca X, Menaud J, Muller G, Lawall J. Entropy: a consolidation manager for clusters. Proceedings of the 2009 ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, Washington, DC, USA, 2009; 41–50.

[29] Feller E, Rilling L, Morin C. Snooze: a scalable and autonomic virtual machine management framework for private clouds. Proceedings of the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Ottawa, Canada, 2012; 482–489.

[30] Verma A, Dasgupta G, Nayak TK, De P, Kothari R. Server workload analysis for power minimization using consolidation. Proceedings of the 2009 USENIX Annual Technical Conference, San Diego, CA, USA, 2009; 28–41.

[31] Speitkamp B, Bichler M. A mathematical programming approach for server consolidation problems in virtualized data centers. IEEE Transactions on Services Computing (TSC) 2010; 3(4):266–278.

[32] Cardosa M, Korupolu M, Singh A. Shares and utilities based power consolidation in virtualized server environments.

Proceedings of the 11th IFIP/IEEE Integrated Network Management (IM), Long Island, NY, USA, 2009; 327–334.

[33] Kumar S, Talwar V, Kumar V, Ranganathan P, Schwan K. vManage: loosely coupled platform and virtualization management in data centers. Proceedings of the 6th International Conference on Autonomic Computing (ICAC), Barcelona, Spain, 2009; 127–136.

[34] Cloud Computing (2010), Wikepedia;en.wikipedia.org/wiki/

[35] Dr. Rich Wolski, (2010) Enterprise Cloud Control.

[36] Ezhil Arasan Babaraj, (2009), Driving Technology Direction on Cloud Computing Platform, Blog post; Hybridfox: Cross of Elasticfox and Imagination, ezhil.syscon.com/.

[37] GoGrid.com. http://gogrid.com/.

[38] IBM. ILOG CPLEX, http://www.ilog.com/products/cplex/

"Smart Bus Alert System for Easy Navigation of Blind"

Adarsh Holikatti *M.Tech student*, **Dr. S. Mohan Kumar** *Associate Professor*
Dept. of Mech., MCE Hassan, Karnataka, India
adiait07@gmail.com , dr.s.m.kumar@gmail.com

ABSTRACT- "There are many techniques which are used for navigating the visually challenged people, navigation in real time traffic is the main problem. Objective of the project is to provide a solution with the aid of wireless sensor networks (WSNs). ZigBee system is used for indicating the presence of blind person in the bus station. Voice module and APR9600 audio playback systems are used to update and inform the blind person about the bus arriving and reaching destinations and to guide him as to what he has to do next. Microcontroller analysis the information provided and generates the corresponding bus number. ZigBee transceiver sends the bus number and announced in the microphone attached with the system. The system is connected with GPS which indicates the destination given. Audio output is generated by the voice synthesizer. The expected outcome of the project is to obtain an easy navigation system for people with visually impaired."

Keywords**: APR9600, ARM7, GPS, Voice module, ZigBee.**

I. INTRODUCTION

Society works essentially through the smooth trade of merchandise, administrations, and brotherhood. Be that as it may, data and assets are made most promptly accessible to the eye. The societal framework and trade system are intended to streamline the opportunity, working, and delight in located individuals - confronting the visually impaired with rejection from this system.

The world is loaded with risks and ponders which society accepts the utilization of vision to maintain a strategic distance from or appreciate. Being visually impaired limits their exposures to these marvels and expands their risk to the perils. More undermining than being cut off from business and societal trade is the contrary condition of general world awareness in regards to visually impaired individuals. Mainstream thinking has dependably fought that visual deficiency drives specifically to lack and insufficiency. Our point is to add to making their lives ordinary in the little way that we can.

As indicated by the measurements and predicts of the WHO upgraded in 2014, 285 million individuals are assessed to be outwardly disabled around the world: 39 million are visually impaired and 246 have low vision. Each outwardly impeded individual countenances diverse difficulties taking into account their particular level of vision. With the ascent of different backing based associations, all the more outwardly disabled individuals have been given the chance to instruction and numerous different means.

However the issues of route for the visually impaired are still exceptionally mind boggling and troublesome particularly when they strolled down in road furthermore explore to inaccessible spots by open transport framework. Blind people might be unwilling to move openly and easily or, out of anxiety, society limits development of the visually impaired person. Deliberate, self-coordinated development is viewed as one of the all the more difficult ranges confronted by visually impaired individuals. While absence of sight is regularly remunerated by improving different faculties, social boundaries and systems of over insurance frequently hamper the perceptual advancement and improvement of useful development in visually impaired individuals.

Guide puppies and strolling sticks take into consideration a free method for navigation, however they are restricted in new situations. RFID is doable and financially savvy however it is more appropriate for indoor correspondence as it were. Likewise it gives stand out way correspondence and a short range of identification. For open air correspondence, all the blind people trust that the guide route offices can manage them like an ordinary individual, and ensure that they are constantly advantageous and safe out and about. The motivation behind this project is to reduce the troubles confronted by blind person when taking city transports, using interactive wireless communication system.

II. MOTIVATION

To use technology for the welfare of the society which includes visually challenged people. The project outcome is indirectly related to the "Digital India" concept which is introduced by Govt. of India. Smart city concept is also in its development stage which aims to bring about change in public transportation system. Thus the project is present day concept and

hence supports innovation of the current/existing system.

III. EXISTING SYSTEM

Consider the case of blind how he confronts the accompanying issues, when utilizing open transport.

Trip arranging – finding a stop/station - finding a passageway to the station - exploring inside the station - finding the right stage and holding up spot - knowing when the right vehicle arrives - finding a vehicle passageway - installment - finding a seat - withdraw on right stop - exploring inside the station - finding the way out of the station - finding the destination
The vast majority of these assignments are paltry for the located, yet exceptionally troublesome for the outwardly debilitated. There are situations when a visually impaired individual has spent a few hours on the transport stop, since he couldn't perceive entry of the right vehicle. What's more, present framework has taking after disadvantages.

- Manual operation
- Monitoring relies upon driver
- Alertness of the system is less
- System is unsafe

IV. PROBLEM STATEMENT

With expansion in movement and populace of the city areas the blind people confront a great deal of obstacles while venturing out starting with one point then onto the next. Because of this most of the blind people are compelled to stay inside and abandon their objectives and dreams as it might include driving from one spot to different spots and in this way costing them their profession/future. This anticipate archives the configuration and usage of a safe visually impaired route framework for the blind people to help them in going from their present area to their craved destination.

V. PROPOSED SYSTEM

To conquer the drawbacks of available accessible assistive gadgets, we propose a Wireless sensor system framework with ZigBee for blind identification in the bus station and installed framework for giving the bus information, lastly GPS for destination sign. Proposed system has following features.

- Safety concerns for blinds
- Automatic operation
- Continuously quick monitor
- High alert system.

VI. SCOPE OF THE PROPOSED SYSTEM

- Blind can undoubtedly get the data about the transport to achieve destination, so travelling makes simple to him.
- Can travel autonomous of any persons need.
- User – friendly interactions with the user.
- Easy to use.
- Audio and vibration alert.
- Voice based input for destination target.
- This is not limited to just visually impaired individual it likewise helps senior individual.
- Communication is given between the visually impaired and driver if there should be an occurrence of any crisis.

VII. BLOCK DIAGRAM

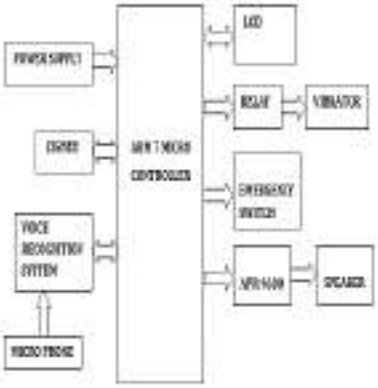


Fig 1: Blind unit.

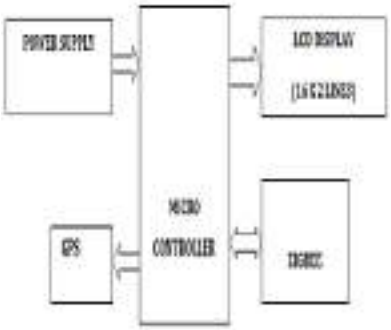


Fig 2: Bus unit.

VIII. WORKING DESCRIPTION

Stage 1: Acquisition of bus arrival information

The ZigBee in the blind module receives the signal which is transmitted by the ZigBee in the bus module within the 30 meters range. So that the blind person can easily get the information about bus arrival.

Stage 2: Intake of the destination to be travelled by the blind person

The blind person gives an audio input of the destination he wishes to reach to the system.

Stage 3: Reception of information by the bus

The blind person gives the input about destination to the voice module V2 and voice module translates the voice of blind person to text and sends it to microcontroller.

Stage 4: Processing of bus information

Once the got signal changed over to text, it should be matched with the destination database present on the bus module so that the framework can illuminate the blind person if that bus is heading off to his wanted destination and in the event that he ought to take that specific bus or wait for the next one.

Stage 5: Audio output for blind interaction

Once the got signal has been decoded, the information is utilized to encourage the sound interface. A voice playback module is interfaced for redesigning the individual about different informations, for example, getting on and off the transport. In this anticipate GPS is utilized for the visually impaired individual to understand that his stop has arrived. At the point when the transport contacts the visually impaired individual's wanted destination he is again educated by his module that he ought to get off the transport.

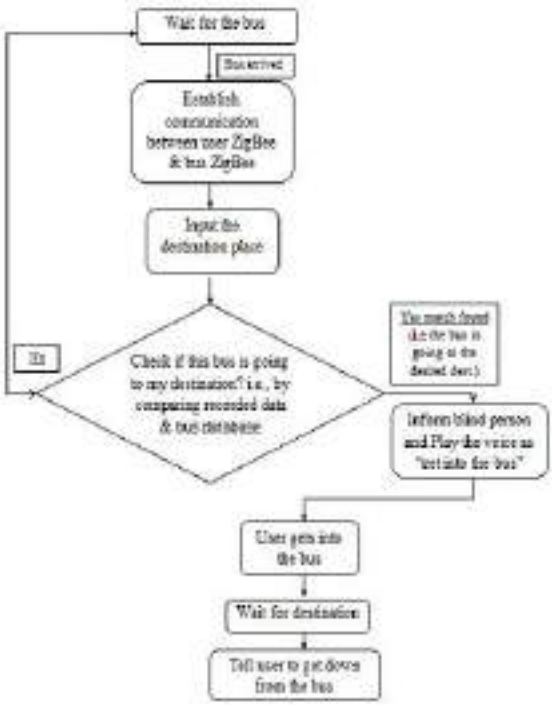


Fig 3: Flow chart.

The above figure demonstrates stream gflow chart. At the point when the blind module is turned ON, the ARM7 controller is initialized. The system waits for the ZigBee to establish the communication. The APR9600 plays that the bus has arrived loudly and also vibration alert for the blind upon the reception of the signal from the ZigBee on bus module. Now the system waits for the blind to give the destination as voice input. All of the above processes are displayed on the LCD

Once the bus module is turned ON it gets initialized and if it is in range of the user module ZigBee then it establishes connection and receives the data that is input from the blind person and the 8051 compares it with the database present on it. It sends back acknowledgement back to the user whether the bus goes to the destination or not.

IX. SYSTEM DESIGN MODEL

A. SOFTWARE DESIGNING:

The modules are required to be programmed for the operation. In this is project, Embedded C programming language is used by utilizing KEIL µvision software. This is popular software that helps in creating embedded C programs, source code editing and debugging, compiling, execution can be done in one single environment. Developed program is dumped into the microcontroller memory by the programmer by the help of Flash Magic software.

B. HARDWARE DESIGN MODULE

Every single operation ought to be controlled. For the control activities, microcontrollers are utilized. ARM7 and 8051 controller are used in the blind and bus module respectively. ARM Board-LPC214X is a breakout board for LPC2148, ARM7TMDI based microcontroller. The LPC2148 microcontrollers depend on a 32-bit ARM7TDMI-S CPU with embedded trace support and real time emulation, which join the microcontroller with installed flash memory. Since the bus module does not require a 32-bit controller as it doesn't have the more number of functions as the blind module, 8051 uC is utilized rather than another arm7 to spare expense without trading off on the usefulness. The Intel 8051 is a 8-bit microcontroller which implies that most accessible operations are constrained to 8 bits. There are 3 essential "sizes" of the 8051: Short, Standard, and Extended. The Short and Standard chips are regularly accessible in DIP (double in-line bundle) structure, yet the Extended 8051 models frequently have an alternate structure figure, and are not "drop-in perfect". To have the communication between the both modules, ZigBees are utilized. ZigBee gives remote RF correspondence. It Works on IEEE 802.15.4 standard and has low power utilization and has a scope of 30 meters which is appropriate for this application. The destination is given as contribution to the arm7 controller through voice module. There are

different voice modules accessible like V1, V2, V3. Contrasted with V1, V2 is easy to control. But just serial information or yield of V1, V2 has other helpful approaches to control and yield the outcome. APR9600 is a sound recorder IC with playback limit for 60 seconds. It is utilized as a part of this anticipate to overhaul and advise the blind person about the bus arriving and achieving destinations and to guide him with reference to what he needs to do next. Vibrator is a small brushless DC motor and it is utilized as a part of this anticipate to give vibration alarm. This is one of the undeniable advantages with cellular telephones, you can get warnings when the gadget is in your pocket without upsetting those around user. GPS permits recording or making areas from spots on the earth and helping user explore to and from those spots. Here it is used to recognize the destination. Global Positioning System (GPS) satellites telecast signals from space that GPS recipients, use to give three-dimensional area (scope, longitude, and elevation) in addition to exact time. GPS receivers gives solid situating, route, and timing administrations to overall users on a ceaseless premise in all climate, day and night, anyplace on or close to the Earth. GPS receivers can secure GPS signals from 65 stations of satellites and yield position information with high exactness in to a great degree testing situations and under poor sign conditions because of its dynamic antenna and high sensitivity. LCD is utilized here to show the operations which are occurring. A LCD is a little minimal effort show. It is easy to interface with a microcontroller due to an installed controller(the dark blob on the back of the board).This controller is standard crosswise over numerous presentations (HD 44780) which implies numerous microcontrollers (counting the Arduino) have libraries that make showing messages as simple as a solitary line of code. Thus it is essential gadget in inserted framework. It offers high adaptability to user as he can show the required information on it.

X. CONCLUSION

With this proposed scheme, a visually impaired person can successfully travel from his location to his desired destination using a bus independently without any hassle.

XI. FUTURE SCOPE

This system can further be improved by using GSM to provide communication between blind and his/her relatives in case of any emergency about more realistic location of his arrival and destination.

XII. RESULTS

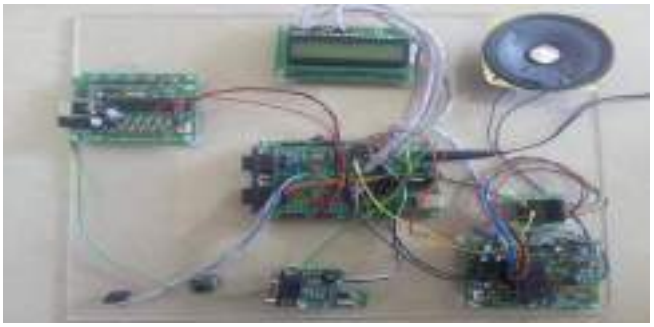


Fig 4: Blind module.

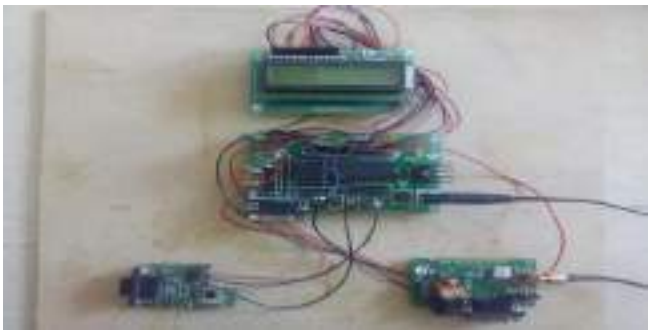


Fig 5: Bus module.

Images shown below are the experimental results.

REFERENCES

[1] Olivier Venard, Geneviève Baudoin, Gérard Uzan, "Field Experimentation of the RAMPE Interactive Auditive Information System for the Mobility of Blind People in Public Transport : Final Evaluation" Intelligent Transport Systems Telecommunications,(ITST),2009 9th International Conference, Oct. 2009.

[2] Syed Rizal Alfam Wan Alwi, Mohamad Noh Ahmad, "Survey on Outdoor Navigation System Needs for Blind People," IEEE Student Conference on Research and Development (SCOReD), 2013, pp.144-148.

[3] Davor Virkes, Karmen Nenadic, Diana Zerec Virkes, "Positioning systems for mobility of the blind, overview and analysis," 46th International Svmoosium Electronics in Marine, ELMAR-2004, 16-18 June 2004.

[4] Zadar, Croatia Passenger BUS Alert System for Easy Navigation of Blind, 2013 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2013]

[5] Mounir Bousbia-Salah, Abdelghani Redjati, Mohamed Fezari, Maamar Bettayeb, "AN ULTRASONIC NAVIGATION SYSTEM FOR BLIND PEOPLE," IEEE International Conference on Signal Processing and Communications (ICSPC 2007), pp.1003-1006.

[6] Jin Liu, Jingbo Liu, Luqiang Xu, Weidong Jin, "Silicon Eyes: GPS-GSM based Navigation Assistant for Visually Impaired using Capacitive Touch Braille Keypad and Smart SMS Facility," IEEE paper 2013.

[7] Mohsin Murad, Abdullah Rehman, Arif Ali Shah, Salim Ullah, Muhammad Fahad, Khawaja M. Yahya, "RFAIDE – An RFID Based Navigation and Object Recognition Assistant for Visually Impaired People," IEEE paper, 2011.

[8] Adrien Brilhault, Slim Kammoun, Olivier Gutierrez, Philippe Truillet, Christophe Jouffrais, "Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning," IEEE paper, 2011.

[9] Bin Ding, Haitao Yuan, Li Jiang, Xiaoning Zang, "The Research on Blind Navigation System Based on RFID."

Audio-video Geners Classification Using AANN

Dr. K. Subashini, Vijayanand.S
Professor, Department of Computer Science and Engineering,
Rajarajeswari college of Engineering, Bangalore-74, India.
Associate Professor ,Department of Computer Science and Engineering,
ACS college of Engineering, Bangalore-74, India.
subashinikrishnaswamy@gmail.com, kgsvanand@gmail.com

Abstract—**In this work presents a method for automatic genres classification into one of sixty-four predefine advertisements of various channels classes used in South Indian TV Broadcasting. Hierarchical Approaches are obtained effective results audio-video classification is very useful to all multimedia retrieval. Features used for categorizing audio are Mel frequency cepstral coefficients. Visual features are extracted using color histogram features in the video clips. The experiments on different Sixty-Four genres illustrate the results of classification done automatically are significant and effective. The results of audio and video confidence score are combined at each level(all six levels) using weighted sum rule for automatic audio-video based genres classification. This method genres classification using AANN systems constructed for sixty four advertisement and results obtained with an accuracy of 98 %.**

Keywords-*Autoassociative neural network (AANN), Mel frequency cepstral coefficients, Color histogram, Audio and video Classification, Audio and video Classification, Weighted sum rul*

I. INTRODUCTION

Last few decays growth of information technology and multimedia information are flooding in the form of audio, video, text and audiovisual. All advertisement broadcasters as well as commercial advertisement broadcasters are enabled with devices to easily broadcast and store multimedia data contents. This data, once huge advertisements are broadcast and stored, are not changed for any case. Manual handling of various advertisement videos are impractical for real time campaigning applications because of its increasingly huge volume of multimedia data.

Hence, it is important to have a method of automatically genres classification of multimedia data for various advertisement videos from the broad casting contents. Categories of audio and video is one important step for automatic indexing and retrieval systems. Our main objective in this paper is to confidence score audio and video classification at all the six levels are combined using weighted sum rule.

II. BACKGROUND

A. Related work

Last few decades, there have been many studies on automatic audio and video classification and segmentation using several features and techniques. In [1], a generic audio classification approach for multimedia classification and retrieval method is described. Unsupervised speaker segmentation with residual phase and MFCC features is given in [2]. The method described in [3] uses content-based audio classification and segmentation by using support vector machines. The work in [4] speech/music segmentation using entropy and dynamism features in a HMM classification framework. The technique described in [5] developed a reference platform for generic audio classification. In [6] audio classification system is proposed using SVM and RBFNN. The perceptual approach is used for automatic music genre classification based on spectral and cepstral features in [7]. A hierarchy based approach for video classification using a tree-based RBF network is described in [8]. In [9] a method is proposed for video classification using normalized information distance. Visual database can be perceptual and categorized into different genres in [10].

The technique described in [11] uses combining multiple evidences for video classification. In [12] the authors address the problem of video genres classification for the five classes with a set of visual features, and SVM is used for classification. Huge literature reports can be obtained for automatic video classification in [13]. Several audio-visual features have been described in [14] for characterizing semantic content in multimedia. The edge based feature, namely, the percentage of edge pixels, is extracted from each key frame for classifying a given sports video into one of the five categories, namely, badminton, soccer, basket ball, tennis and figure skating techniques in [15]. A feature, called motion texture, is derived from motion field between video frames, either in optical flow field or in motion vector field in [16]. In [17] GMM is used to model low level audio/video feature for the classification of five different categories namely, sports, cartoon, news, commercial, and music. An average correct classification rate of 86.5% is achieved with one hour of records per genre, consisting of continuous sequences of five minutes each and 40 second decision window. Combining the evidence obtained from several complementary classifiers can improve performance based on the literature shown in [18] and [19]. In [20] a survey of audio based music classification and annotation is described. Then, in [21] a survey on visual content based video indexing and retrieval shows huge information on video. A effective algorithm for unsupervised speaker segmentation using AANN is described in [2]. In [22] a robust speaker change detection algorithm is proposed. Evaluation of indexing techniques for audio indexing is described in [23]. In [24] a hybrid approach is presented for audio

segmentation. Acoustic, strategies for automatic segmentation are described in [25]. In [26] unsupervised speaker change detection using SVM misclassification rate is described. Automatic segmentation, classification and clustering of broadcast news audio is given in [27].

B. Outline of the work

In these work two systems (audio and video) based on two(audio and video) modeling techniques (AANN) approach is used optimal class obtained boundary between the classes by learning from training audio and video data. Results of sixty four advertisement genres are classified obtain using AAN center approximates a cluster of training data vectors (audio and video) such that they are close to each other in calculating Euclidean space. Here a vector is input to the AANN, the centers that are very near to that particular vector become strongly activated, in turn activating certain output nodes used in five layers AANN architecture. Experimental results show that the two systems (audio and visual) genres classification better results obtain using hierarchical approach to get maximum accuracy of through AAN.

The paper is organized as follows: Feature extraction is presented in Section III. Modeling techniques used for classification and hierarchical approach described in Section IV. Experimental results and conclusions described in Section V, and VI, respectively.

III. FEATURE EXTRACTION FOR CLASSIFICATIONS

a. Acoustic Feature Extraction for Classifications

MFCC is perceptually motivated representation defined as the cepstrum of a windowed short-time signal. A non-linear mel-frequency scale is used which approximates the behavior of the auditory system. The MFCC is based on the extraction of the signal energy with-in critical frequency bands by means of a series of triangular filters. Whose centre frequencies are spaced according to melscale. The mel-cepstrum exploits auditory principles as well as the de-correlating property of the cepstrum [2][31][32]. Fig. 1, Illustrates the computation of MFCC features for a segment of audio signal which is described as Follows: The mel-frequency cepstrum has proven to be highly effective in recognizing structure of music signals and in modeling the subjective pitch and frequency content of audio signals. Psychophysical studies have found the phenomena of the mel pitch scale and the critical band, and the frequency scale-warping to the mel scale has led to the cepstrum domain representation.The mel scale is defined as

$$F_{mel} = \frac{c \log\left(1 + \frac{f}{c}\right)}{\log(2)}$$

Where F_{mel} is the logarithmic scale of f normal frequency scale. The mel- cepstral features can be illustrated by the MFCCs, which are computed from the fast Fourier transform (FFT) power coefficients. The power coefficients are filtered by a triangular band pass filter bank. When c in (1) is in the range of 250 - 350, the number of triangular filters that fall in the frequency range 200 - 1200 Hz (i.e., the frequency range of dominant audio information) is higher than the other values of c. Therefore, it is efficient to set the value of c in that range for calculating MFCCs. Denoting the output of the filter bank by S_k (k = 1, 2, · · · ,K), the MFCCs are calculated as

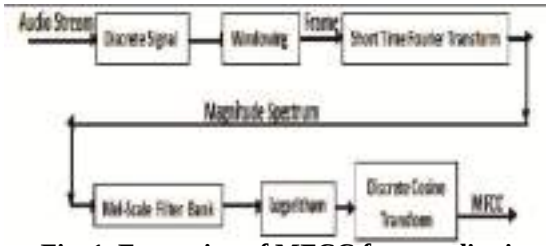


Fig. 1. Extraction of MFCC from audio signal

3

MFCCs are short-term spectral features as described above and are widely used in the area of audio and speech processing. To obtain MFCCs [2],[31],and [32]., the audio signals were segmented and windowed into short frames of n samples. Magnitude spectrum was computed for each of these frames using fast Fourier transform (FFT) and converted into a set of mel scale filter bank outputs.

Logarithm was applied to the filter bank outputs followed by discrete cosine transformation to obtain the MFCCs. For each audio signal we arrived at 39 features. This number, 39, is computed from the length of the parameterized static vector 13, plus the delta coefficients 13, plus the acceleration coefficients

b. Visual Feature Extraction for Classificationtions

A color histogram is a representation about distribution of colors in a representation about distribution of colors in an image, derived by counting the number of pixels in each of the given set of color ranges in a typically two dimensional (2D) color space. A histogram of an image is produced first by dicretization of the colors in the image into a number of bins, and counting the number of image pixels in each bin. This is described in audio-video based segmentation and classification using SVM and AANN [31][32].

The histogram provides a compact summarization of the distribution of data in a image. The color histogram of an image is relatively invariant with transaction and rotation about the viewing axis, and may vary very slowly with the view angle. Further, they are computationally trivial to compute. Moreover, small changes in camera viewpoint has on color histograms. Hence, they are used to compare images in many applications. This work uses color histogram as visual feature. The RGB color space is quantized into 64 bins by n.

Fig.2 Color Histogram

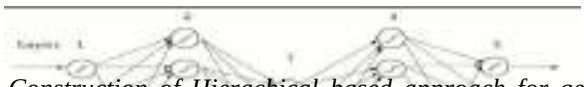
The RGB (888) color space is quantized into 64 colors. For each frame 320*240 size, there are 64-dimensional feature vector are extracted. In order to reduce the dimension of the feature vector, only the dominant top 16 values are taken as features in experiments and is shown in Fig. 2.

IV. MODELING TECHNIQUES USED FOR CLASSIFICATION

a. Autoassociate Neural Network (AANN)

Autoassociative neural network models are feed forward neural networks performing an identity mapping. The AANN is used to capture the distribution of the input data [29], [30] and [31]. Let us consider the five layer AANN model shown in Fig. 6, which has three hidden layers. The processing units in the first and third hidden layers are non-linear, and the units in the second compression/hidden layer can be linear or non-linear. As the error between the actual and the desired output vectors is minimized, the cluster of points in the input space determines the shape of the hyper surface obtained by the projection onto the lower dimensional space. A five layer auto associative neural network model is used to capture the distribution of the feature vectors. The second and fourth layers of the network have more units than the input layer.

The third layer has fewer units than the first or fifth. The activation functions at the second, third and fourth layers are non-linear. The non-linear output function for each unit is tanh(s), Where s is the activation value of the unit. The standard back propagation learning algorithm is used to adjust the weights of the network to minimize the mean square error for each feature vector. The AANN captures the distribution of the input data depending on the constraints imposed by the structure of the network, just as the number of mixtures and Gaussian functions do in the case of Gaussian mixture model.



B Construction of Hierachical based approach for geners classification using AANN: Here we are using only two models for hierarchical classification systems are constructed as shown in Fig.3., The audio based hierarchical classification system uses AANN and MFCC features and the video based classification system uses AANN and color histogram features.

Fig. 3 A five layer AANN Model

Each box in the hierarchical classification system represents an AANN with the structure 39L 68N 14N 68N 39L for audio and 64L 128N 18N 128N 64L for video. In this structure there are 2^i auto associative neural networks in level i. For example in the last level n there are auto associative neural networks. These networks are trained using 2^n different languages sports data. In the $n^{th,}$, $n\text{-}1^{th}$ level, the features of two n^{th} level AANNs are combined to create an AANN as shown in Fig.4.

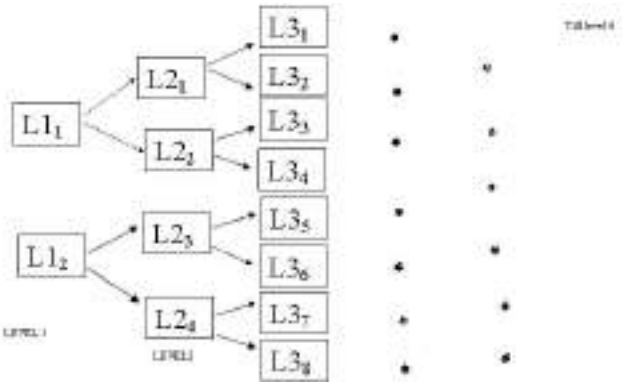


Fig.4 Hierarchical indexing system

b. Audio and Video Classification using AANN

Auto associative neural network is used to capture the distribution of the acoustic and visual feature vectors of a category. Separate AANN model are trained to capture the distribution of acoustic and visual feature vectors of each category. In testing process we used every acoustic and visual feature vector is given as input to each of the models. The output of the model is compared with the input to compute the normalized squared error. The normalized squared error is transformed into a confidence score as described in Section 4.2. Average value of confidence score is calculated for all models. The category is decided based on the maximum confidence score. This is described in audio-video based classification using AANN [31][32].

$$m_j = \frac{w}{n} a_j + \frac{1-w}{p} v_j, 1 \le j \le c,$$

Where $a_j = \sum_{i=1}^{n} x_i^j \;\; v_j = \sum_{i=1}^{p} y_i^j$

c_i^a - Category label for i^{th} audio frame.

c_i^v - Category label for i^{th} video frame.

vj - video based score for j^{th} category.
aj - audio based score for j^{th} category.
mj - Combined audio and video based score for j^{th} category.
c - number of categories.
n - number of audio frames.
p - number of video frames.
w - weight

The category is decided based on the maximum mj .

Similarly, the results obtained for audio and video classification by AANN are combined using:

$$s = \frac{w}{n}\sum_{i=1}^{n} s_i^a + \frac{(1-w)}{p}\sum_{i=1}^{p} s_i^v$$

Where n - number of frames in audio signal.

p - Number of frames in video signal.

s_i^a - Confidence score rate of the i[th] audio frame.

s_i^v - Confidence score of the i[th] video frame.

s - Combined audio and video confidence score.

w - Weight.

The category are decided based on the maximum confidence score obtained from the models.

V. EXPERMENTAL RESULTS

In our experiments, 64 advertisements of TV broadcasting video are recorded with a resolution of 320*240 pixels and at 8 KHz with 16-bits per sample. The LPCC, LPC, and MFCC features are extracted as comparing the features results we optimal result obtain in MFCC that is described in section 3.1 and similarly edge feature, motion features and color histogram features are extracted but overall performance of color histogram obtained effectively that is described in section 3.2.

For conducting experiments, audio and video data are recorded using a TV tuner card from various televisions south Indian language channel at different timings to ensure quality and quantity of data stream. The training data test includes various duration it can be 2, 4, 6-mins of audio stream for each genres duration various such that 2, 4, 6-mins of video stream for each genres. Audio stream is recorded at 8 KHz with mono channel and 16 bits per sample. Video clips are recorded with a frame resolution of 320×240 pixels and frame rate of 25 frames per second. Training data is segmented into fixed overlapping frames (in our experiments we used 160 ms frames with 80ms overlapping). The sample features extraction process for repeated for audio and video data of varying durations.

A six level audio classification system is created using AANN and MFCC features. Similarly, a six level Video classification system is created using AANN and color histogram features. The number of each AANN in each level is 2, 4, 6, 8, 32 and 64 respectively.

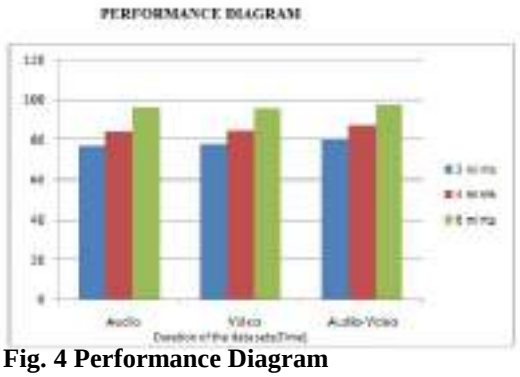


Fig. 4 Performance Diagram

For classifying a advertisement video from a test clip, the MFCC and color histogram features are extracted for the test clip and are given as input to the first level auto associative neural networks. The output of the model (O) is compared with the input to compute the normalized squared error (e_k) for the test feature vector Z is given by [31][32]

$$e_k = \frac{\|z - o\|^2}{\|z\|^2}$$

Where o is the output vector given by the model. The error e_k is transformed into a confidence score s using [31][32].

$S = exp(-e_k)$

The confidence score from the audio system and video system are combined using a weighted sum rule and the highest confidence score at each level is used to find the classification path in the genres classification system. This process is repeated for all the levels and the required various advertisement video is classified in the sixth level. The performance diagram of 100 sample of each advertisement duration various 2 mints, 4 mints and 6 mints are shown in Fig.4. Experimental results show that the proposed method achieves a classification rate of about 98.0%.

VI. CONCLUSIONS

This paper proposed a method for genres classification using AANN. Audio and video based hierarchical classification system was obtained using sixty four advertisements. Audio based hierarchical genres classification system was constructed using MFCC features and video based hierarchical classification system was constructed using color histogram features. Auto associative neural network models are used to capture the distribution of these features. The performance of the system was evaluated using confidence score of audio and video data sets and using the weighted sum rule (Proposed method) achieves a classification rate of about 98.0%.

REFERENCES

[1] S. Kiranyaz, A. F. Qureshi, M. Gabbouj, A generic audio classification and segmentation approach for multimedia classification and retrieval, IEEE Trans. Audio, Speech and Lang Processing 14(3)(2006) 1062–1081.

[2] S. Jothilaskmi, S. Palanivel, V. Ramalingam, Unsupervised speaker segmentation with residual phase and MFCC features, Expert System With Applications 36 (2009) 9799–9804.

[3] L. Lu, H.-J. Zhang, S. Z. Li, Content-based audio classification and segmentation by using support vector machines, Springer-Verlag Multimedia Systems 8 (2003) 482–492.

[4] J. Ajmera, I. McCowan, H. ourlard, Speech /music segmentation using entropy and dynamism features in a HMM classification framework, Speech Communication 40 (3) (2003) 351–363.

[5] R. Jarina, M. Paralici, M. Kuba, J. Olajec, A. Lukan, M. Dzurek, Development of reference platform for generic audio classification development of reference plat from for generic audio classification.,

IEEE Computer society, Work shop on Image Analysis for Multimedia Interactive, (2008) 239–242.

[6] P.Dhanalakshimi, S.Palanivel, V.Ramaligam, Classification of audio signals using SVM and RBFNN, Expert System With Applications 36 (2009) 6069– 6075.

[7] C. Lin, J. Shih, K. Yn, H. Lin, Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features, IEEE Transactions Multimedia 11 (4) (2009) 670–682.

[8] W. Gillespie, D. Nguyen, Video classification sing a tree-based RBF network, IEEE International Conferance on image processing 3 (1) (2005) 465–468.

[9] K. Kaabneh, A. Abdullah, A. Al-Halalemah, Video classification using normalized information distance., In Proceedings of the geometric modeling and imaging-new trends (2005) 34–40.

[10] M. Geetha, S.Palanivel, V.Ramaligam, A novel block intensity code for video classification and retrieval, Expert System With Applications 36 (2009) 6415–6420.

[11] V. Suresh, C. K. Mohan, R. Kumaraswamy, B. Yegnanarayana, Combining multiple evidence for video classification, In IEEE international conference intelligent sensing and information processing (2005) 187–192.

[12] V. Suresh, C. K. Mohan, R. Kumaraswamy, B. Yegnanarayana, Content-based video classifica tion using SVM., In International conference on neural information processing.

[13] D. Brezeale, D. J.Cook, Automatic video classification a survey of the literature, IEEE Transaction on System, Man, and cybernetic 38 (3) (2008) 416–430.

[14] Y. Wang, Z. Liu, J. Huang, Multimedia content analysis using both audio and visual clues, IEEE Signal Process. Mag. 17 (2000) 12–36.

[15] Y. Yuan, C. Wan, The application of edge features in automatic sports genre classification., In Proceedings of IEEE Conference on Cybernetics and Intelligent Systems (2004) 1133–1136.

[16] Y.-F. Ma, H.-J. Zhang, Motion pattern based video classification using support vector machines.,In Proceedings of IEEE International Symposium on Circuit and Systems 2 (2002) 69–72.

[17] L. Q. Xu, Y. Li, Video classification using spactial-temporal features and PCA, International Conference on Multimedia and Expo 3 (2003) 345–348.

[18] J. Kittler, M. Hatef, R. Duin, J.Matas, On combining classifier, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[19] L. Xu, A. Krzyzak, C. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Trans. Syst. Man, Cybern. 2 (1992) 418–435.

[20] Z. Fu, G. Lu, K. M. Ting, D. Zhang, A survey of audio-based music classification and annotation, IEEE Transactions Multimedia 13 (2) (2011) 303– 318.

[21] H. V. Weiming, N. xie, L. Li, X. L. Zeng, S. maybank, A survey on visual content-based video classification and retrival, IEEE Transaction on System, Man, and cybernetic part c (2011) 1–23.

[22] J. Ajmera, I. McCowan, H. Bourland, Robust speaker change detection, IEEE Journal of Signal Process Letter 11 (8) (2004) 649–651.

[23] J. A. Arias, J. Pinquier, R. Ande-Obrecht, Evaluation of classification techniques for audio classification, In proc. 13th Eropean conf. Signal Processing.

[24] S. Cheng, H. Wang, Metric SEQDAC: A hybrid approach for audio segmentation, Proc. 8th International conference on spoken language Process. (2004) 1617–1620.

[25] T. Kemp, M. Schmidt, M. Westphal, A. Waibel, Acoustic strategies for automatic segmentation of audio data, Proc. IEEE International conference on Acoust, Speech, Signal Process. (2000) 1423– 1426.

[26] P. Lin, J. Wang, J. Wang, H. Sung, Unsupervised speaker change detection using SVM training misclasssification rate,IEEE Int'l Conf. Acoustics, Speech and Signal Processing 14 (3) (2006) 1062– 1081.

[27] M. Sieglar, U. Jain, B. Raj, R. Stern, Automatic segmentation, classification and clustering of broadcast news audio, Proc. DARPA Speech recognition workshop (1997) 97–99.

[28] V. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.

[29] S. Palanivel, Person authentication using speech, face and visual speech, Ph.D thesis, Indian Institute of Technology Madras, Department of Computer Science and Engg (2004).

[30] B. Yegnanarayana, S. Kishore, AANN: An alternative to GMM for pattern recognition, Neural Networks 15.

[31] K.Subashini,S. Palanivel, V. Ramaligam., Audio-video based segmentation and classification using SVM and AANN. International Journal of Computer Applications, **53(18) ,PP. 0975-8887, Sep 2012.**

[32] K.Subashini,S. Palanivel, V. Ramaligam., Audio-video based segmentation and classification using AANN. **International Journal of Computer Application and technology, Volume:1, Issue:2,PP 53-56, Sep-Oct 2012.**

A SURVEY ON IMPORTANCE AND APPLICATIONS OF DATA MINING USING ARTIFICIAL NEURAL NETWORK

Ms. R.Jaya, Mr.Giridhar V
Sr. Assistant Professor, Department of Computer Science and Engineering
U G Student, Department of Computer Science and Engineering,
New Horizon College of Engineering, Bangalore, Karnataka, India
jayamanojkumar@gmail.com, giridharv10@gmail.com

ABSTRACT-As there have been increased modernisation curve in development field of each and every equipment, process or technique involved in modern medicine world. In Data mining pattern recognition and pattern extraction are interpreted, explored, realised and analysed to finally evaluate, arriving at expected conclusion having uplifting accurate and respective diagnosis to respective cause. DM techniques with utmost focus on NN and ANN as well as DM tasks of conceptual inference in mining and matching are imprinted in this paper

Keywords - DM techniques, DM tasks, ANN, NN, Association rules, Naive bayes, Clustering

I. INTRODUCTION

An artificial intelligence mechanism exercising computational practices exploring patterns to predict essence of recognized patterns in particular field for a particular cause with a brief analysis from statistically stored database catalogues leads to current domain studies in data mining, adaptive flourishing domain of computer science in medical field. The normal lifestyle, adulterate food, habits and habitats etc., has given verdict to reduced lifespan. Data mining has built up statistical analysis of exaggeration and success, building the domain from the scratch with new conceptualization of disease cure.

1.1 Importance of data mining: Data mining is core knowledge from where essential data is captured or mined for some basis .its gaining demand in Business transactions, Scientific data analysis, Medical research and personal data scrutiny, Surveillance video and pictures queries and also in field of Satellite sensing, Games, Virtual Worlds. We extract the required data in a substantial manner from databases indulging in cleaning database, integrating elements of designated data to form data warehouses further selecting and transforming to task related data mining involving pattern extraction to gain knowledge[16][9].

Data mining is an interdisciplinary subfield in computer science domain with computational process of discovering patterns in big data involving collective methods of artificial intelligence, machine learning, statistics, analytics etc., Data mining involved with fetching data from enormous warehouses imply task querying called data mining task primitives, these primitives form interactive communication during discovery in order to direct the mining process, or examine the mining deeds from different angles or depths.

1.2 Related research work

World towards advancement has been improving technology in every domain but still lifespan of humans is not meeting up to the mark as it's expected to be, according to bio science. The present lifestyle, threatening to be the cause of reduced lifespan. There has been series of heart, kidney, and lung etc., disorders being encountered [1] echo's the diagnosis of heart disease using predictive DM where the survey report of heart patients is collected and diagnosis is involved with Naïve Bayes and Decision Tree algorithms. The prediction of disease is by DM weka tool where dataset containing attributes and values forms input and heart patients prediction forms output. The brief discussion about dataset, tools and principles involved in predictive DM is emphasized.

The naïve bayes algorithm has imprinted a new channel of exploration in understanding data[6]. Conveys the principle, implementation, objectives, rules and step by step algorithm analysis of naïve algorithm along with Jelinek-mercer smoothing technique[2]. Intelligent heart disease prediction system mechanism envelopes analysing, predicting heart diseases based on person's lifestyle and symptoms he suffers from. The system which is expandable incorporating numerous records or attributes with significant rules under DM, also extracts masked essential historic data from medical database. jelinek smoothing algorithm has more weighed thrust in prediction answering complex queries with a flexible move, providing ease of access detailed information and accuracy[7].

Technology dependency on bio system or bio system on technology is incredible. Introducing chemical kinetics of life, modelling kinetics of drug release, classification and species determination is made through advanced computational technologies as ANN, briefly discussing the neural network model with mathematical analysis showing network learning and database training structures, made clinical context as input for ANN to respond giving out diagnosis related output, explaining overviewed picture of cancer, cardiovascular and diabetes diseases. Finally describing fundamental steps in ANNs-based medical diagnosis involving features such as -Building the database Data cleaning and preprocessing Data homoscedasticity -Training and verification of database using ANN Network type and architecture Training algorithm Verification Robustness of ANN-based approaches -Testing in medical practice etc., are depicted[8].

The improvement in field of science is has made us to march towards discovering new diseases classifying them to get a cure. Artificial intelligence based expert system gobble up entire catalogue database of crucial cases critically, to provide taxonomy depicting essential description. An expert system (ES) known as knowledge based system, is a computer program that uses knowledge and inference procedures to solve problems that are ordinarily solved through human expertise. It has mainly three components a) knowledge base, b) inference engine c) user interface. The diagnosis of coronary heart diseases (CHD) is dealt along with proposed framework of algorithmic descriptions and rule set of CHD such as Angina pectoris, Myocardial infarction(heart attack) etc., are focused[9].

Renewed interest in the 2000s in Neural Networks was sparked because of the advent of deep learning. In CMOS, Computational devices have been created, to both neuromorphic computing and biophysical simulation. In the computational devices had big improvement from 2006 [11].

In this paper [14] the author tells us that in data mining techniques neural network is widely used. Multilayer perception is very slow for the high dimensional datasets. In real time data mining the application it is an important draw back for using them, because its requirement is that, it needs answer within a short duration of time [13]. CMAC Neural Network adaptation they have proposed for data mining in this research paper. Which provide guaranteed convergence and fast training time? This paper also defines us that how CMAC adaptation is constructed for data mining, about classification model that is used in data set of real-life. It works on real dataset. The models which were used are CMAC, MLP and C4.5 and the algorithm is CMAC algorithm [13].

In this paper [15] the author tells us about a new method in which DSS is supported based on data mining and neural network. Three subjects have been discussed in this paper. To support DSS how to make use of neural network is the first. And the second one is that, to support DSS how to make use of data mining and the third one is, To support DSS how to make use of neural network and Data Mining. DSS was developed by American scientist Scott

Three departments of DSS are discussed they are Knowledge system (KS), and Problem process system (PPS) and language system (LS). The neural network to DSS function is ensuing by scrutinizing the characteristic of problems and neural network of DSS. The large scale parallel distribution processing function, learning function, selflearning of natural language processing system , overcoming the difficulties of infinite recursion and assembled blast, promoting of deciding the ability of DSS and real time processing , Adoptive parallel association reasoning [15].

II. CATEGORIES OF DATA MINING

Data mining tasks are of two categories: Descriptive data mining and predictive data mining, where descriptive DM describes information of interest without predetermined analysis. Predictive DM analyse to fetch result in predictive manner by allowing unknown field values to be submitted, followed with interpreting patterns encountering output information with pattern mapping experience.

DM categories involve classification, prediction, clustering and association rules, the former two categories descend from prediction DM and latter categories descend from description DM[6][9].

Classification: The primary step under focus in DM is to classify by projecting different items to their respective groups through pattern recognition skill and experienced examination. Classification is supervised learning as data classes or concepts are predefined describing even the characteristics of a set of data classes.

Prediction is of construction, estimation, perception, detection along with effective utilisation of predicted one's. it tracks the class of unlabeled object or needed value from varied range frequency, followed with forecasting to generate the result of accuracy by estimating future value of continuous variables based on live patterns within the study.

Clustering is unsupervised learning process which envelope classification techniques, with no group predefinition. The key concept of clustering is to group abstract object based on quantitative similarities. Clustering is not convergent to specific algorithm but its general and divergent[6][16].

Association rules are being used in research which aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories having communicative effective interface associating matching protocols[1][16].

2.1 Neural Networks: The inspiring traits for neural networks being animal behaviour responding to stimuli. It's one of the best algorithms in prediction and diagnosis of illness. Cognitive science based NN, An artificial intelligence based data mining technique being made reputation in healthcare technology.

The input information is perceived by dendrites of neuron generating a non linear response when adequate threshold information is reached and collected at synapse to communicate and pass other neurons via axon. All attributes values must be encoded in standardized manner .This technique works well as long as minimum and maximum values are known and all new potential data at reference are in boundary. NN are robust to minor disturbances if violations exceed control certain ad hoc solutions or extensions of boundary will be undertaken.

NN algorithm: it works through training data set neurons record by record adjusting weights constantly reducing prediction errors. It involves training through number of passes rendering conclusion at required diagnosis. Training period if reduced affects collapsing the models efficacy. As NN are prone to over fittings as they flexible enough causing threshold to reach less than minimum criteria. The impact of abrupt termination is catastrophic; we avoid such termination by cross validation check procedure involving five steps.

1. To retain part of the original data set as a holdout validation set.
2. Proceeding to train neural network as above on remaining training data.

3. Apply the weights learned from the training data on the validation data
4. Monitor two sets of weights, one current set of weights produced by training data, and one best set of weights as measured by the lowest Sum Squared Error so far on the validation data.
5 When the current set of weights has significantly greater SSE than the best set of weights

2.2 ANN's: A mathematical representation of human neural architecture, reflecting its learn content and skill full knowledge with generalisation capabilities. It belongs to artificial intelligence domain, imprinting research related models, echoing high level non-linear mechanism patching complex relationships among the variables describing threshold weight to data mine critical information to be popped out at required times. Moreover, ANN is an adaptive system that changes structure of threshold weighted communicative neural mechanism based on inputs that enter during learning phase of training.

A neural network is giant clusters of numerous neurons which are framed to fit in an organized layer view. The weighted connection confirms the neural sensitive data transfer approach to frame ANN methodology. The structure of neural network is formed in between layers, having a layer of inputs giving raise to next set of hidden layers to march out at output layer.

ANNs are at more fame by their accuracy, has made footstep in replacing traditional diagnosis and classification approaches. they are nonlinear, nonparametric, adaptive and can theoretically approximate any fundamental relationship with arbitrary accuracy [3].

Choosing ANNs in diagnosis has its importance for these generalised reasons as an iterative training approach involved in analysing weighted solutions ,simple architecture for physical implementations, the feature of generalising input vectors, complex class distribution mapping ,ANNS ability to be implied as arbitrary function approximation mechanism that grasps from observed data. ANN deals with function approximation problem which involves selection of a specific function among a well-defined class to closely match a target function in a task-specific way.

2.3 Working of ANN:

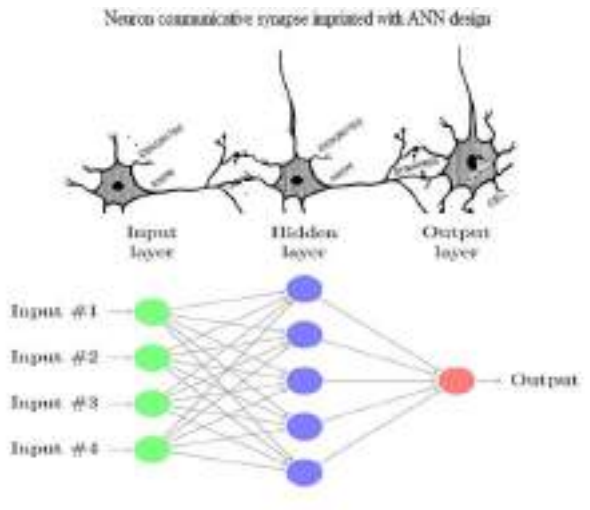


Fig 2: Neuron communicative synapse imprinted with ANN design

Reticulate structure of intricate neurons clustering dynamically, interactively in a self organised way to process the information of interest.' Integrated circuit and advanced technology with two dimensional devices having limited layered interconnections 'of fact are brought to physical reality considering restriction to types, scope of artificial neural networks while implementing with silicon. Neural networks of current medicinal technology are simple clustered web of primitive artificial neurons .layered clustering mechanism forming intricate network to resolve biomedical problems in optimised procedure is thoroughly through engineered human toil of glide.

Generalised similar neural network layered structure or topology for many medical cases has been employed. The topology has some neuron interface to experience real circumstance in fetching inputs .few other neurons provide the real world with the network's outputs. The output being the representation of input, handling the reflex arc, the output view has scanned mechanism of patternising the image of importance generating response to stimuli. The sandwiched neural layers between outer and inner layers are hidden to sight.

Earlier research was made connecting neurons at random manner, which remain false positive towards success. Even the primitive invertebrates considering from mollusc phylum have structured neural ganglia or brain, then enhanced research incorporating layered network approach, summation and transfer function involved promoting to compute in computer functional neural networks procedure.

The lines of communication during transactions between neurons form the thesis of neural network interactive approach. The package wrapping selfdom communication between layers form the asset of neural networks, implying variable input strength of neural networks giving out result of necessity. There are two communicative behaviour of connections one being summing information to neighbouring neuron layer called "excite behaviour" and another one being subtractive information of no advancement to next neuron called "inhibit behaviour". The probability of pattern match should yield approximate result ceasing confusion leading to important case called lateral inhibition. Neurons of same layers when dealing with probabilities avoid confusions called lateral inhibition. The network chooses the highest probability inhibiting all others using lateral inhibition. This conceptualisation of lateral inhibition is also called competition.

Connections are of feed forward pathway and feedback pathway, the former approach is forward communication considering from input to output layer and in the latter approach the output of one layer routes back to a previous layer. The way that the neurons are connected to each other has a significant impact on the operation of the network. The implementation of professional software package to schedule controls such as add, delete etc., coordinates to excite or inhibit behaviour as employed in neural study.

The first step towards Working of ANN is signed with training ANN, the structure for an application once prepared to be functioned then mechanism is triggered towards training. The loads of training expressed with varied weighed analysed input note or data are chosen randomly at initial circumstance, followed with learning and training.

There are two approaches to training - supervised and unsupervised.

In supervised training both inputs and outputs are decided which are provided for processing. The network then processes the inputs and compares its resulting outputs against the optimised expected or desired output. If errors or exceptions discovered the corrective procedure is triggered to back propagate all through the system re-correcting encountered flaws in communicative mechanism. Back propagation has been tedious, as it needs to locate capturing point of flaw and correct it by adjusting the weights of input which control the network. This process is being iterative as it's involved with adjusting varied range of weights, allowing the neural data of answered result to flow. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights are ever refined. The current commercial network development packages provide tools to monitor how well an artificial neural network is converging on the ability to predict the right answer focusing optimal view. These tools allow the training process to be continuously active for some days until the system reaches some statistically desired point of accuracy predicted.

To monitor the network to determine if the system is simply memorizing its data in some non significant way, supervised training needs to hold back a set of data to be used to test the system after it has undergone its training. New adaptive training techniques are fed based on necessity. When finally the system has been correctly trained, and no further learning is needed, the weights can, if desired, be "frozen."

Unsupervised training: the model is not provided exactly the ordered sequence of inputs and outputs; the procedure followed during training is not step by step. The network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self organization or adaption. This adaption to the environment is the promise which would enable science fiction types of robots to continually learn on their own as they encounter new situations and new environments. This unexpected aspect to life and the human desire to be prepared, there continues to be research into, and hope for, this field. Based on statistical characteristics or properties only we cluster input data in classes. Cluster significance and labelling is employed. The labelling can be carried out even if the labels are only available for a small number of objects, representatives of the desired classes[2][3].

2.4 Application of neural networks*:* Function approximation, regression analysis, time series prediction, Call control, sequence recognition, novelty detection, sequential decision making, Data processing, Language Processing, Character Recognition, Image (data) Compression, Pattern Recognition, Signal Processing, Servo Control ,financial sector database etc.,[4]

2.5 Application of artificial neural networks: Load Forecasting, Fault Diagnosis\Fault Location , Economic Dispatch, Security Assessment, Alarm Processing, Eddy current analysis, Harmonic source monitoring, Load frequency control (Automatic Generation Control), Hydroelectric Generation Scheduling, Power System Stabilizer Design ,medical diagnosis, financial application[5].

III. CONCLUSION

At present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. The commercial, educational and scientific applications are increasingly dependent on these methodologies. This paper has provided the summary of one of data mining technique ANN used for medical data mining in the field of healthcare and also the diseases they classified. It also expresses the importance and challenges of data mining with respect to neural and artificial neural networks.

REFERENCES

[1] Sotiris Kotsiantis, Dimitris Kanellopoulos "*Association Rules Mining: A Recent Overview*" Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece

[2] http://www.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf

[3] Application of data mining using artificial neural network :survey", *Intl journal of database theory and application,vol 8 2015*

[4] http://www.psych.utoronto.ca/users/reingold/courses/ai/cache/neural3.html

[5] Sonali Maind,Priyanka Wankar, "Research paper on basic of artifical neural netwok ",*Intl Journal of recent and innovation trends in computing and communication vol2* issue 1 jan2014

[6] R. Jaya, S. Mohan Kumar "A Study On Data Mining Techniques, Methods, Tools And Applications In Various Industries" *Int J Current Research Rev | Vol 8* • Issue 4 • February 2016

[7] Filippo Amato, Alberto López,et al "Artificial neural networks in medical diagnosis ",*Journal of applied biomedicine 2013* issn 1214-0287

[8] R.A. Soltan M. Z. Rasha, B.El-Desouky, "Diagnosis of Some Diseases in Medicine via computerized Experts System" vol 5,october 2013 *International Journal of computer science and information technology*

[9] R. Jaya, S. Mohan Kumar" Ayurveda Medicine Roles in Healthcare Medicine, and Ayurveda towards Ayurinformatics" *IJCSMC, Vol. 4*, Issue. 12, December 2015, pg.35 – 43

[10] Nikita Jain , Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER " IJRET: *International Journal of Research in Engineering and Technology* eISSN: 2319-1163 | ISSN: 2321-7308 Volume: 02 Issue: 11 | Nov-2013

[11] Rasika P. Ghom , N. R. Chopde" Survey Paper on Data Mining Using Neural Network" *International Journal of Science and Research (IJSR)* ISSN (Online): 2319-7064

[12] Mariammal,Dayanthi,Dr,Patra, "Major diease diagnosis and treatment suggestion system using data mining techniques" , *IJARCST 2014,VOL2 JAN-MACH 2014*

[13] F. Palacios, X. Li and L. E. Rocha, "Data mining based on CMAC neural networks", Electrical and Electronics Engineering, 2006 3rd International Conference, IEEE, (2006), pp. 1-4.

[14] T. Nagabhushan, H. Ko, J. Park, S. Padma and Y. Nijagunarya, "Classification of Symbolic Objects Using Adaptive Auto-Configuring RBF Neural Networks", Information Technology Convergence, 2007. ISITC 2007. International Symposium IEEE, (2007), pp. 22-26

[15] X. Qian and X. Wang, "A New Study of DSS Based on Neural Network and Data Mining", E-Business and Information System Security, 2009. EBISS'09. International Conference IEEE, (2009), pp. 1-4

[16] R. Jaya, S. Mohan Kumar "A Survey On Medical Data Mining Healthcare Related Research And Challenges*"* *International Journal of Current Research Vol. 8*, Issue, 01, pp.25170-25173, January, 2016

R.JAYA is working as a Sr.Assistant Professor in the department of CSE at New Horizon college of Engg.She has 13 years of teaching experience. She has done bachelor and master degree in engineering under Anna University. She is doing Ph.D under VTU. Her research interest includes data mining, expert systems,SOA and security.

Methodology To Make Natural Language as Computer Programming Language

[1]Nalina H J, [2] Mr.Vinay T R
Department of Computer Science and Engineering, VTU, Bengaluru-562157
nalina0309@gmail.com, tr.vinay@gmail.com

ABSTRACT-Natural Language Processing(NLP) holds great promise for making computer interaction easier with the naïve users or non-computer programmers. We propose a methodology which converts natural language statements into a computer programming language. The non-computer programmers to express their ideas in computer language (programming language) is difficult. So, here we propose a methodology where the naïve users can easily express their ideas in their natural language on computers. We can either build a natural language compiler but, this takes time and effort for developing and learning. So our methodology helps the naïve users to simply express their idea in their own natural way and our process converts it into a computer program and generates the output.

Keywords: Natural Language, Computer Language, Naïve users, Computer Programmers, Natural language programming.

1. Introduction

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction[1]. Many challenges in NLP involve: natural language understanding, enabling computers to derive meaning from human or natural language input; and others involve natural language generation.

Modern NLP algorithms are based on machine learning, especially statistical machine learning. The paradigm of machine learning is different from that of most prior attempts at language processing. Prior implementations of language-processing tasks typically involved the direct hand coding of large sets of rules. The machine-learning paradigm calls instead for using general learning algorithms often, although not always, grounded in statistical inference to automatically learn such rules through the analysis of large corpora of typical real-world examples. A corpus is a set of documents that have been hand-annotated with the correct values to be learned[2].

Many different classes of machine learning algorithms have been applied to NLP tasks. These algorithms take as input a large set of "features" that are generated from the input data. Some of the earliest-used algorithms, such as decision trees, produced systems of hard if-then rules similar to the systems of hand-written rules that were then common. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input feature[3]. Such models have the advantage that they can express the relative certainty of many different possible answers rather than only one, producing more reliable results when such a model is included as a component of a larger system.

The main difference between the Natural Language and the Computer Language should be known in order to understand the Natural Language Processing.
Natural Language Processing (NLP) and Programming Languages are both established areas in the field of Computer Science, each of them with a long research tradition. Natural Language Processing involves a wide range of techniques that enable the automated parsing and processing of natural language[4]. This paper tries to address this gap by proposing a methodology that attempts to convert natural language text into computer programs. This methodology is particularly usefull for non-computer programmers to express their ideas on computer[5].

Natural Language	Computer Language
1)It is generally peoples uses to share their feeling or for communication purpose. 2)It is understood and developed naturally by humans. 3)Used to communicate between humans. 4)Very Strict syntax is not followed. 5) The Lifespan is endless.	1)It is a derived from Natural Language which is used to instruct machines. 2) It is developed by humans in the form of programming language to coomunicate with machines. 3) Used to communicate with the machines and human interactions. 4)Very Strict syntax has to follow. 5)The Lifespan is limited.

Table 1: Natural language v/s Computer language

NLP is a huge field in computer science about language based interactions between computers and humans. For any person to express the ideas,he has to learn computer programming language. All most all compilers today are built using English like language. So all the prospective programmers should know English language[6]. For non-English persons, he has to learn English language then learn a computer programming language. So to help non-English prospective programmers we have to build a natural language compiler for their language or use a convertion tool, which can convert natural language statements into programming statements which is executable[7].

2. Existing system

Since, its inception in 1950s, NLP research has been focusing on tasks such as machine translation, information retrieval,

text summarization, question answering, information extraction, topic modeling, and more recently, opinion mining. Most NLP research carried out in the early days focused on syntax analysis.

Although the semantic problems and needs of NLP were clear from the very beginning, the strategy adopted by the research community was to tackle syntax first, for the more direct applicability of machine learning techniques. Another popular model for the description of natural language is production rule. A production rule system keeps a working memory of on-going memory assertions[8].

There are no tools available in the market to convert the natural language into a programming language but tools are available which convert natural language words into its equivalent English words. Nowadays only the computer programmers are able to develop or implement the ideas. The naïve users are just using the computer systems but they are not able to develop or express their ideas on computer. As they don't know the programming languages[9]. Also the naïve users must depend on the computer programmers to implement their ideas.

3. Proposed System

The proposed system explains to converting methodology for moving from naturally speaking language to machine understandable language. To develop a natural language compiler it takes time and effort. Also we have to train the people as how to write computer programs using natural language compiler. Here we do proposing a methodology which will convert a natural language statements into a computer language statements[10]. And this methodology can be further applied to any natural language.

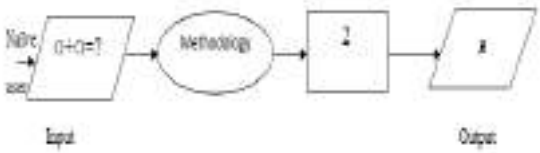


Fig 1 : High level design of Proposed Methodology

The above figure 1 describes the processing steps towards from natural to programming language. The user or the naïve user who will give the input in the form of natural language statements as ೦+೦=? which will be given to our process, where the computer program will be executed in order to make the input understable by computer. After the process is done inside the methodology the output will be generated in the english language, and again that will be converted into the natural language output.

1.Algorithm

These steps gives a processing operation from input to executable output.

[1] Given statements is broken into a meaningfull phrases(Lexical Analysis).

[2] Phrases is compared with the stop words.

[3] Phrases are analysed to parts of speech.

[4] Operation to be performed with respect to input statements will be analysed.

[5] Computer program is executed.

[6] Output will be generated.

[7] Later the generated output will be converted back to the natural language.

Case Study: The following statements is taken for case study.

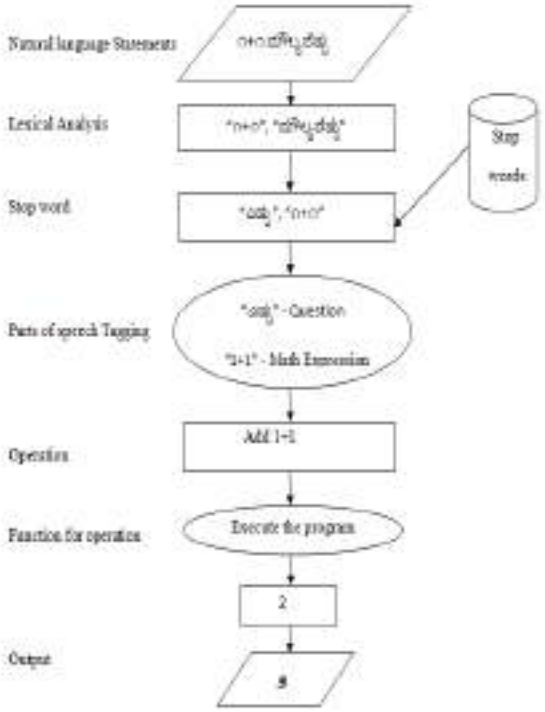


Fig 2 : A Simple Case Study

Fig 2 shows an example program that asks for a addition of 2 numbers in Kannada. Here the first statement is "೦+೦ ಮೌಲ್ಯವೆಷ್ಟು " which is a Natural Language statement. This statement is later broken into meaning small pieces called tokens which we call Lexical Analysis of the given statement "೦,+೦,", "ಮೌಲ್ಯವೆಷ್ಟು". Next the word "ಮೌಲ್ಯವೆಷ್ಟು" is reduced to "ಎಷ್ಟು"using stop words. Now ಎಷ್ಟು " is analysed as questioning and ೦ is converted into 1. So ೦+೦ is converted into mathematical expression 1+1. The operation will be performed to generate the output through the computer program as "2". The last step is to convert back the above generated output(2)into natural language as (**2**).

Starting with the natural language text as an input we would ideally like to generate a computer program. While this is still a long term goal, in this section we show how we can automatically generate computer program skeletons that can

be used as a starting point for creating procedural computer programs. Specifically, we focus on the description of a system for natural language procedural programming: Mathematical expression.

4. Conclusion

This methodology is developed in order to make a naïve users to express their ideas on computer. This may lead to bust in the number of programmers and many new novel ideas may be developed. This will help all the naive users to program and develop their ideas. Hence due to this the world of technology becomes easier. As there is no necessary to learn the programming language to program the ideas. So the common people can also work in the technology field by quickly adopting.

REFERENCE

[1] NLP (Natural Language Processing) for NLP (Natural Language Programming) by Rada Mihalcea, Hugo Liu, and Henry Lieberman. Springer-Verlag Berlin Heidelberg 2010.

[2] From NLP (Natural Language Processing) to MLP (Machine Language Processing) by Peter Teu,Udo Payer and Guenter Lackner. Volume 3685 of Lecture notes in computer science. Springer (2012).

[3] Jumping NLP Curves: A Review of Natural Language Processing Research by Erik Cambria, Bebo White. IEEE Computational intelligence magazine | may 2014.

[4] Natural Language Processing Laboratory Plan by Mohammad Azadnia, Sina Rezagholizadeh and Alireza Yari. December 2009.

[5] TechWare: Spoken Language Understanding Resources by Gokhan Tur, Ye-Yi Wang and Dilek Hakkani-Tür. IEEE SIGNAL PROCESSING MAGAZINE May 2013.

[6] TAN Ying-wei,MO Qian. Design and Realization of Supervise Adaptive Topic Tracking System. Journal of Zhengzhou University (Natural Science Edition). 2014.

[7] Pang B,Lee L,Vaithyanathan S.Thumbs up? Sentiment classification using machine learning techniques. In: Isabelle P,ed. Proc. of the EMNLP 2002. Morristown: ACL,2010.

[8] Ye Shiren, T-S Chua,M-Y Kan, and L. Qiu, "Document concept lattice for text understanding and summarization," Inf. Process. Manage., vol. 43, 2014.

[9] S. Pado and M. Lapata, "Dependency-based construction of semantic space models," Assoc. Comput. Linguist., vol. 33, no. 2,2014.

[10] R. Keefer and N. Bourbakis, "Interactions with a mobile reading device for the visually impaired," in Proc. 21st Int. Conf. Tools with Artif. Intel., Nov. 2012.

A Novel Frame Work for E-Ndorse

Divya R K, Bhanu K N
Deportment of MCA, Rajarajeswari College of Engineering, Bangalore, India
divya.karthigayan@gmail.com, bhanu.kn@gmail.com

ABSTRACT-The proposed framework is used to develop system where various business enterprises and individuals can advertise their products and services. The purpose is to provide a centralized repository of information, which can be easily maintained and accessed.

Index Terms— Endorse, Classifieds, Ad - Advertise.

I. INTRODUCTION

The proposed system is designed in such a way it is possible to advertise products and services with ease and minimum expenses. The system allows various interested parties to register and advertise their products, which can be viewed by a large number of users and customers. It is the web application supports various features and functionalities to publish and manage the advertisements.

In existing system for advertising purpose big hoardings, which is costly and within few days that will be torned out or shaded. To advertise again the person or the organization has to invest or spend lot of money. So to overcome this disadvantage the system named E-Ndorse is proposed. Main Problems in Existing system developing one stop solution for maintenance of online advertisements. Ads Manager Plus is a Business to Business application which facilitates wide range of businesses to provide classifieds online.

The organization of paper is as follows: Section.1 explains the architecture. Section.2 explains about E-Ndorse as a system, Section.3 explains about System Design, Section.4 explains about Implementation and Section.5 concludes the paper.

II SYSTEM ARCHITECTURE

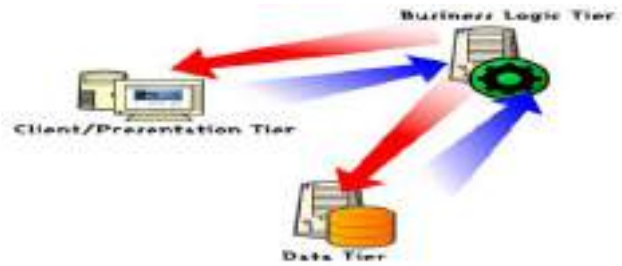


Fig.1 3Tier Architectural Flow

current application is being developed by taking the 3-tier architecture. The 3-tier architecture is the most common approach used for web applications today. In Fig.1 the web browser acts as the client, IIS handles the business logic, and third tier MS-SQL Server handles database functions as mentioned in [3]. Although the 3-tier approach increases scalability and introduces a separation of business logic from the display and database layers, it does not truly separate the application into specialized, functional layers. For prototype or simple web applications, the 3-tier architecture may be sufficient. However, with complex demands placed on web applications, a 3-tiered approach falls short in several key areas, including flexibility and scalability. These shortcomings occur mainly because the business logic tier is still too broad- it has too many functions grouped into one tier that could be separated out into a finer grained model. The proposed system can be designed perfectly with the three tier model, as all layers are perfectly getting set as part of the project. In the future, while expanding the system, in order to implement integration touch points and to provide enhanced user interfaces, the n-tier architecture can be used. The following diagram will represent the typical n-tier architecture.

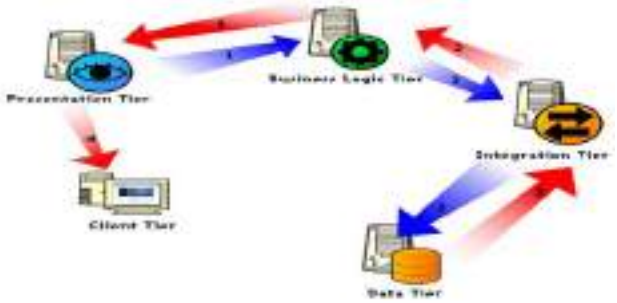


Fig.2.Detailed Flow

III E-Ndorse

The system interface has been developed by keeping the graphics concept in mind that associated through a browsers interface. The GUI'S have been categorized as

A. Administrative user interface

B. The operational or user interface

The administrative user interface concentrates on the consistent information that is practical, part of the organizational activities and which needs proper authentication for the data collection. The interfaces help the administrations with all the transactional states like data insertion, data deletion and data updating.

The overall flow is first organization or an individual should login or sign in, then registering the ad of their product and ad is posted based on the category. This System is explained using modules which work as follows: first is Ad Registration module: This module provides a facility for registration of new advertisements of organization or an advertise of an Individual .In this module Organization or an individual registering their ad for posting.

Fig.3 Feature of system

Usually the products are posted based on the category. There are many different categories like generally men category, women category, Babies category and Kitchen, electric items etc. These are called as Parent Category. Under this category there are sub categories like clothes, toys, shoes, mobiles etc. These is maintained and controlled by admin in another module that is:

Category management: In this module all the advertisements and services are maintained under different categories. The number and type of categories will be decided by the administrator of this application.

During registration many organizations will be registered and posted their ads. Single organization can post many products. All such advertise are arranged under the category. Distribution of products according to category of a particular organization.
classifieds: this module maintains the details of all classifieds and advertisements put by a particular individual or organizations. even this classified is maintained by admin only in module called classified administration**:** this module is responsible for maintaining all classifieds, details such as type of classifieds, duration of the classified, validity of the classified, its categories, who is the ownership of that particular classified and others. For example: If any classifieds Due date is closed admin have the authority to delete the post, which is present in respective classified. In this module admin can add new classified.

In this application the front end technology is Java as mentioned in [1]. And backend is Mysql as mentioned in [6] used to store the information of all modules like login module, category module, classified module etc.And jsp is used for coding as mentioned in [5].

IV SYSTEM DESIGN

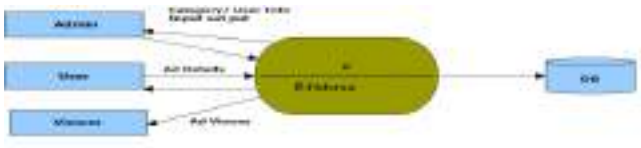


Fig 4 Context Level diagram

Fig.4 explains overall flow. First user adds the details of oraganisation as well as advertisement product details and classified details.Admin has all authority on the all of these information.Admin maintains the ,User information that is organisation or an individual information.finally Viewer can view the postings,can buy and use the services provided.

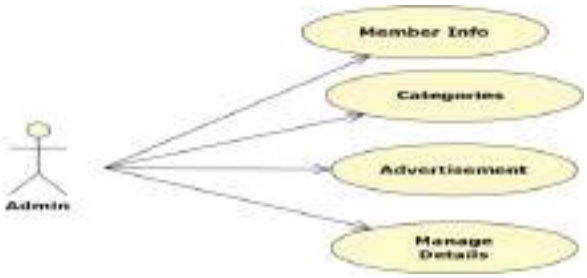


Fig.5 Use Case of Admin

This fig.5 explains what are all the responsibility of Admin manages the member Information, manages the Category, manages the Advertisements and Details all Classifieds. Both Admin and Member work together to form the Application.

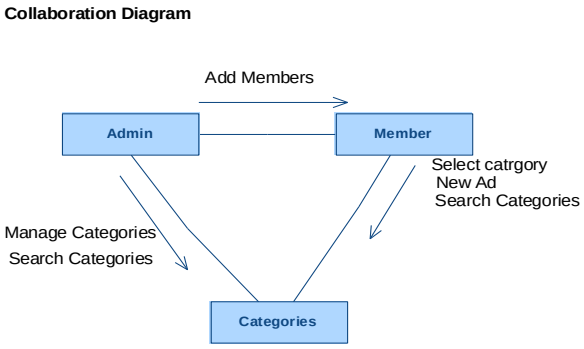


Fig.6 Collaboration Diagram

If no admin management of the application becomes very difficult. Because the admin is the person one who administrates and responsible for all. Similarly if no member no use of endorsement because they are the person who use this system for advertising their products. Admin and Member are the functional requirements. Category is also important so that the products can be arranged in an proper way.

V. IMPLEMENTATION

Fig.7 Ad Registration

Fig.7 Explains ad registration module in which Individual member or organization should register first in case of new user. If exsisting user means no need of registration.without

registration adds can be posted directly by logging in.During registration the member id will be provided for each and every individual and organization.This id should be provided during login, new add posting and mentioning classifieds.So registration is mandatory for new user.

Fig.8 Explains a part of admin module.Admin is the person who administrates and amnages the whole system.

Fig.8 Admin view

Once admin logsin, the admin page will be opened in which few links are provided like member, category and classifieds.In member admin can view the registered members, can add new members.In category number of categories of this system is decieded by the Administrator.Administrator can view, add, delete, and edit the categories. These Administrator, classified, category are the functional requirements of this system.

Fig.9 New add Posting

Fig.9 Describes new add posting.After registration it will be redirected to next form, That is NewAdd Posting form. In that mandatory fields are add id, member id, category id , Product name, Description about product, posted date, closing date, Actual price, offer price and services provided. These informations should be entered.once entering all the informationadd will be posted by clicking the button post add.once it is submitted add is Posted succesfully.

Fig.10 Add details

Fig.10 showing the details of advertisements.This is for admin view, when admin wants to check the posted adds.In this form admin can delete if stock is closed or services are closed or date is closed.Here fields shown are product name, parent category, description of add and price.

VI. CONCLUSION

This frame work helps to develop an application for advertising industries. This can be easily maintained and accessed in low cost compared to hoardings. The application supports various features and functionalities to publish and manage the advertisements industries. This frame work provides a centralized repository of information, which can be easily maintained and accessed.

ACKNOWLEDGMENT

The authors are thankful for the encouragement and support received throughout the work to Management RRGI, Bengaluru and principal RRCE, Bengaluru.

REFERENCES

[1]Herbert Schildt,"Java the Complete Reference"9th Edition.

[2] Mohr Yasin, B. S. L. (2008). Web advertising and attitude: Internet users' view. The Business Review.

[3] Paul Du Bois, Mysql reference, 5th edition.

[4]McGraw-Hill,"JSP Reference"2001.

[5]William Crawford," Java Servlet Programming" 2nd Edition.

[6] www.mysql.com

[7] java/j2ee application framework.

Divya R K is currently PG student of Rajarajeswari College of Engineering, pursuing final year of Master of Computer Application.

Implementation of Inpainting Algorithm

Rohan Ramakrishna Genne, Bhanu K. N
Department of MCA, Rajarajeswari College Of Engineering, Bangalore-74
gennerohan@gmail.com, bhanu.kn@gmail.com

ABSTRACT--The inpainting algorithm is related to image. First a low-resolution image is built from original picture. An inpainting algorithm is applied to fill-in the holes of low-resolution image. The quality of the inpainted image is improved by using a single image super resolution method.
Keywords - Diffusion-Based Method, Exemplar-Based Method, Single Image Super-Resolution Method.

1. INTRODUCTION

With all programming languages, you compile or interpret a program so that you can run it on your computer. The Java programming language is not usual in that a program is both compiled and interpreted. With the compiler, first, you convert a program into an intermediate language called Java bytecode. the platform-independent codes interpreted by the converted on the Java platform. The interpreter parses and runs each Java bytecode instruction on the computer. Compile happens just once. conversion occurs each time the program is executed.

In the implementation of inpainting algorithm first, an input image we have to take. Then in the image, we have to crop some area. The cropped area will be anything in the image. By the inpainting algorithm, the holes are filled.but the quality of the image quality is not good. For the good quality of image single image super-resolution method is used. And at last we can save that image is our system only. It will take less time compared to existing approaches. There are two existing approaches are present. first is Diffusion-based method. By Diffusion-based method, the image holes are not filled correctly and the quality of the image is not good. It looks a blur image. The second approach is the Exemplar-based approach. In Exemplar-based approach, by the frames the image is clear. But it takes long time and frames are also increased. but the proposed system has two main components inpainting algorithm and super resolution method. By this two main components, the image looks like a real image.

2.SYSTEM ARCHITECTURE

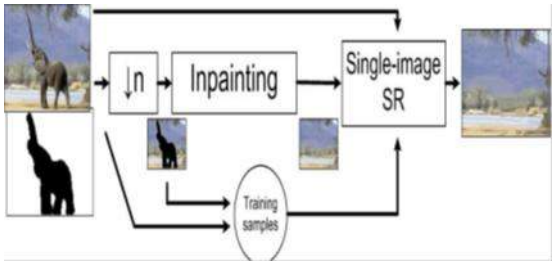


Fig. 1 System Architecture

1.1 Image inpainting

Inpainting is the process of reforming lost parts of images. In the digital world, in painting refers to the application of algorithms to replace lost parts of the image.

1.2 Image Restoration

Image Restoration is the operation of taking a region of an image and forming the clean original image. Corruption may come in many forms such as blur or camera miss focus.

1.3Super-resolution

Super-resolution is a technique that enhances the resolution of the imaging system. In some SR techniques termed optical Super-Resolution the diffraction limit of systems is transcended.

1.4 Detailed Design

1.4.1 Use Case Diagram

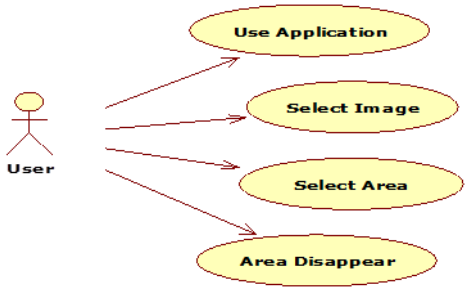


Fig. 2 Use Case Diagram

1.4.2 Sequence Diagram

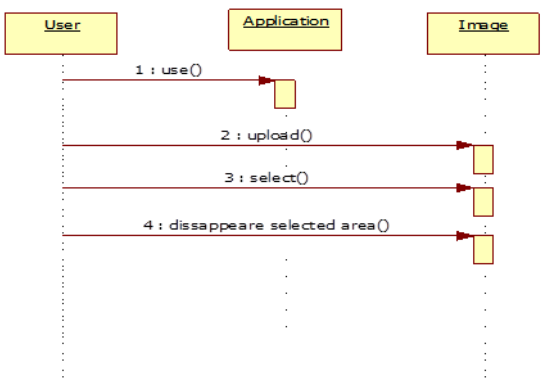


Fig. 3 Sequencial Diagram

3.THE INPAINTING ALGORITHM

Super-resolution is a technique that uses the resolution of the imaging system. There are different super-resolution techniques are present. Like optical single-resolution and geometric single-resolution.The geometric single-resolution is related to the resolution of an image. Images may contain textures. but by the sampling theorem the contents can be automatically restored. For the case of missing areas exact reconstruction is recommended. The inpainting model is successful for the large image the region must be small. If the region is small the result will be more clear. In this algorithm, the human visual system can tolerate some amounting of blurring areas.

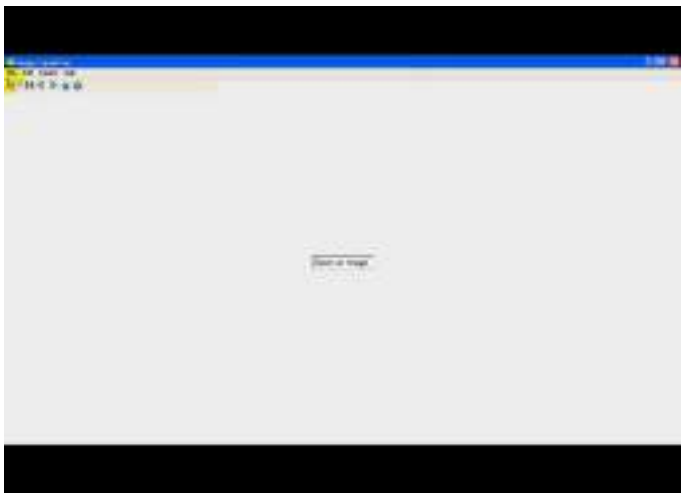
Let Ω be a small area to be inpainted and let ∂Ω be its boundary. since Ω is small, the inpainting by the diffusion process that ∂Ω into Ω. An algorithm reconnects edges reaching to ∂Ω.

This algorithm consists of initializing Ω.by clearing the color information and region to be inpainted with a diffusion kernel. ∂Ω is a one-pixel boundary and the no. of iterations are controlled for inpainting domain by checking none of the pixels belonging to the domain their values are changed more than a certain threshold during the previous iteration.The diffusion process is iterated,the inpainting progress from ∂Ω to Ω.

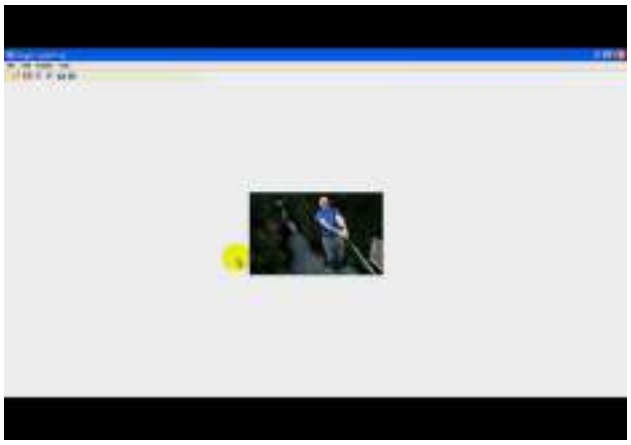
4. IMPLEMENTATION

Module 1 – Image upload

The First module is used to image upload. The image will be taken by our system only.

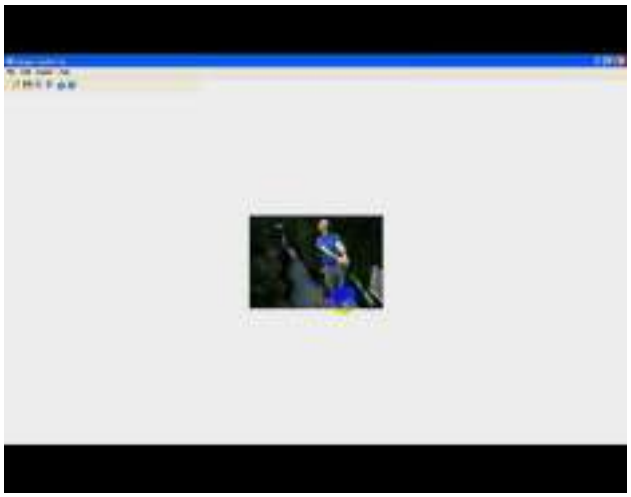


After image upload, the image is displayed.

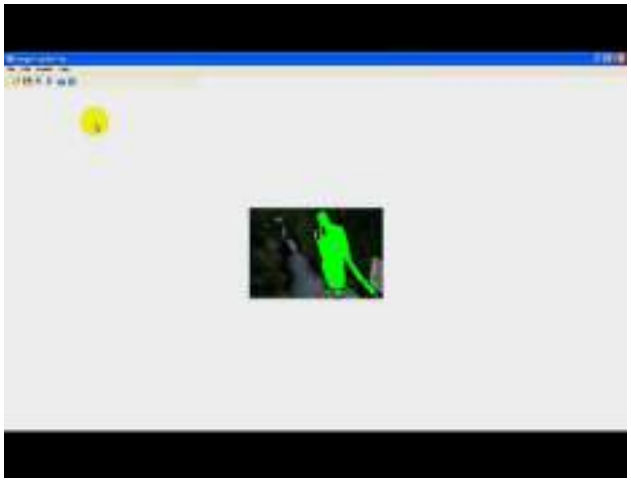


Module 2 – In the Image Cropping Some Area

After image uploading, some area in the image will be cropped.

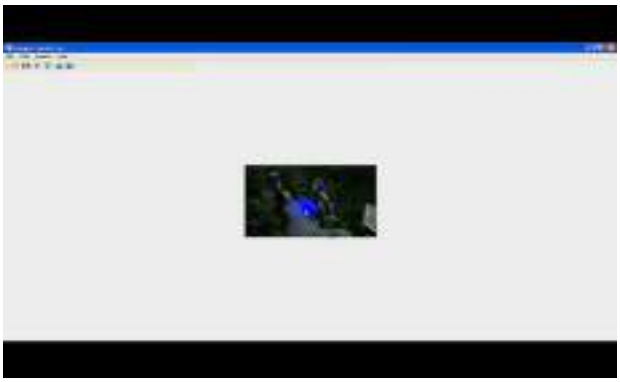


It displays the cropped area in image



Module 3 – Final Output and Saving In System

By inpainting algorithm, this is the final output.



The final output is stored in our system.

5.CONCLUSION

In this project, I have introduced a new inpainting framework. I first propose an extension of a well known exemplar-based method and compare it to existing methods. Then , a super-resolution method is used to recover high-resolution. This is interested for different reason. One interesting avenue of future work is to perform several inpainting of the low-resolution images. First, different kinds of inpainting methods is used to fill the missing areas of a low-resolution image. Second, by given inpainting method, one can fill the missing areas by using different setting. The use of the proposed framework is reduce the computational time.

ACKNOWLEDGMENT

The authors of this paper are thankful to the Management, RRGI, Bangalore and Principal, RRCE, Bangalore for their continuous encouragement and support throughout the work

REFERENCES

Journal Papers:

[1] Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: SIG-GRAPH 2000. (2000)

[2] Tschumperl´e, D., Deriche, R.: Vector-valued image regularization with PDE's: a common framework for different applications. IEEE Trans. on PAMI 27 (2005) 506–517

[3]Chan, T., Shen, J.: Variational restoration of non-flat image features: models and algorithms. SIAM J. Appl. Math. 61 (2001) 1338–1361

[4] Criminisi, A., P´erez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. On Image Processing 13 (2004) 1200–1212

[5] Drori, I., Cohen-Or, D., Yeshurun, H.: Fragment-based image completion. ACM Trans. Graph. 22 (2003) 303–312

[6]Harrison, P.: A non-hierarchical procedure for re-synthesis of complex texture. In: Proc. Int. Conf. Central Europe Comp. Graphics, Visual, and Comp. Vision. (2001)

Genne Rohan Ramakrishna is currently as PG student of Rajarajeswari College of Engineering pursuing final year of Master of Computer Application.

Bhanu K N is currently working as Associate Professor in Department of Computer Applications , Rajarajeswari College of Engineering, Bengaluru. Her research and professional career spans for about 12 years of Teaching and Industrial Experience. She obtained her MCA., from IGNOU, New Delhi, M.Phil., in Computer Science from VMU, Salem and presently Pursuing Ph.D., in Computer Science from Rayalaseema University, AP. Her expertise is primarily in the domains of Wireless Sensor Networks, sensor clouds, Adhoc Networks and Sensor Big data analytics.

A Novel approach of 4*4 Vedic Multiplier using Reversible Logic Gates

Shivarathnamma, G.Jyothi, M Z Kurian
Dept. of ECE, Sri Siddhartha Institute of Technology
Tumkur, Karnataka, India
shivarathnaasha@gmail.com, grandhejyothi@gmail.com, mzkurianvc@yahoo.com

ABSTRACT-The performance of the multiplier determines the system performance, because the multiplier is slowest element in the system. Multiplication is one of the major operation in arithmetic and logical operations and multiplier is used in many applications like FFT, DFT, Image enhancement, DWT etc,. The multiplication speed influences the processor speed, so the speed of the multiplication should be high. There is one such promising solution i.e., Vedic multiplier. Vedic multiplier is designed using Vedic mathematics. Vedic mathematics is an ancient system of mathematics, which is formulated by Sri Jagadguru Bharathi Krishna tirthaji (1884-1960). The word Vedic is obtained from the word "Veda" which gives the meaning of power house of knowledge and devine. This paper proposes the novel approach of 4*4 Vedic multiplier using reversible logic gates. Design of high speed Vedic multiplier is designed based on Vedic mathematics Vedic mathematic. Usage of reversible logic gates leads to reduction of power dissipation. Power dissipation is an important factor which can't be neglected in VLSI. The multiplier is designed using "Urdhva Tiryakbhyam" sutra from Vedic mathematics which is different from conventional multiplier like array and booth multiplier. The coding is done using VHDL for 2*2 Vedic multiplier and simulation is done using Xilinx 14.5 tool. The logic verification of the modules has been done by using Modelsim.

Keywords: Vedic mathematics, Vedic multiplier, Urdhva Tiryakbhyam, Reversible logic gates, garbage outputs, constant inputs, quantum cost, gate counts.

I. INTRODUCTION

Multiplier has many applications like FFT, DFT and image enhancement. In order to increase the operation speed of FFT, DFT the multiplication process should be faster. Compared to basic multiplier like Wallace and array multiplier Vedic multiplier is faster and it consumes less power. The entire speed and the performance of the system is depends on the speed of the addition and multiplication taking place in the system. Delay will be more due to the long multiplication process, and the propagation delay will also be considered because of the parallel adders used in the addition stage.The multiplication done in threeways: Partial Products Generation (PPG), Partial Products Addition (PPA) and Final Convectional Addition. The main issue is to increase the speed of the multiplier partialproducts should be reduce. Vedic mathematics is an ancient system of mathematics, which formulated by Sri Jagadguru Bharathi Krishna tirthaji (1884-1960). The word Vedic is obtained from the word "Veda" which gives the meaning of power house of knowledge and devine. These are 16 sutras in Vedic mathematics Urdhva Tiryakbhyam and Nikhilam Navatascaramam Dasatah are traditionally taken to design multiplier. Vedic multiplier does not produce any partial products so the number of adders required in the multiplier is less which leads to simplicity of the circuit. There are 16 sutras in Vedic mathematics and 13 upa sutras (sub formulae). Out of 16 sutras Urdhva Tiryakbhyam sutra is used to design the Vedic multiplier.

1.1 Reversible logic:

Reversible logic has its basics from thermodynamics of information processing. According to this, traditional irreversible circuits generate heat due to the loss of information during computation. In order to avoid this information loss the conventional circuits are modelled using reversible logic. In 1961 Mr. R. Landauer showed that the circuits designed using irreversible elements dissipate heat due to the loss of information bits [1]. It is proved that the loss of one bit of information results in dissipation of KT*ln2 joules of heat energy where K is the Boltzmann constant and T is the temperature at which the operation is performed. Then in 1973 Mr.C.H. Benett showed that this heat dissipation due to information loss can be avoided if the circuitis designed using reversible logic gates [2]. A gate is considered to be reversible only iffor each and every input there is a unique output assignment. Hence there is a one to one mapping between the input and output vectors. A reversible logic gate is an n –input, n- output device indicating that it has same number of inputs and outputs. A circuit that is built from reversible gates is known as reversible logic circuit.

While designing the circuit using reversible logic gates the designer should concentrate on the following parameters:

1. Garbage Outputs (GO): Indicates the unused outputs in the circuit which cannot be avoided and it is necessary to achieve reversibility. The circuit should contain minimum number of garbage outputs.

2. Constant inputs (CI): This refers to the number of inputs that are to be maintained constant at either 0 or 1 in order to synthesize the given logical function.

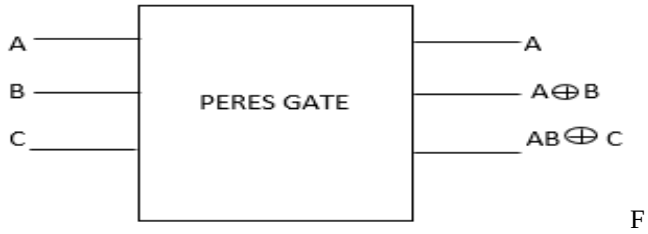
3. Gate Counts (GC): Indicates the number of reversible gates that are required to design the circuit, so that the circuit should contain reversible gates as minimum as possible.

4. Quantum cost (QC): This refers to the cost of the circuit in terms of the cost of a primitive gate. It is calculated knowing the number of primitive reversible logic gates (1*1 or 2*2) required to realize the circuit.

Some of the reversible logic gates are shown below that are used in the proposed design:

1.1.1 PERES GATE:

It is a 3*3 reversible gate i.e., it has three inputs and three outputs. The representation of Peres gate is shown below. Quantum cost of this gate is 4.



F

ig.1: Peres gate.

Peres gate is one of the popular gate and used in many applications.

1.1.2 TOFFOLI GATE:

It is a 3*3 reversible gate i.e., it has three inputs and three outputs. The representation of Toffile gate is shown below. Quantum cost of this gate is 5.

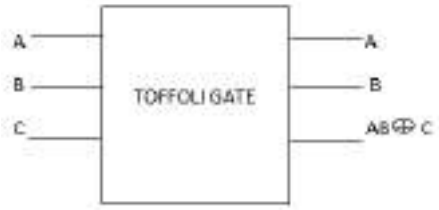


Fig.2: Toffoli gate.

1.1.3 BME GATE:

It is a 4*4 reversible gate i.e., it has four inputs and four outputs. The representation of BME gate is shown below. The quantum cost of this gate is 6.

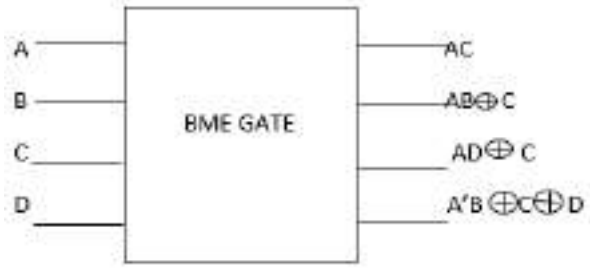


Fig.3: BME gate.

II URDHVA TIRYAKBHYAM MULTIPLICATION ALGORITHM

The Sanskrit word Urdhva Tiryakbhyam which gives a meaning of vertical and crosswise in English. Vedic mathematics and design using Urdhva Tiryakbhyam sutra. This sutra can be apply to Binary, Hex and Decimal multiplication and it generates all partial products concurrently. Fig. shows the 4*4 Binary multiplication using Urdhva Tiryakbhyam sutra and can also be used for N*N bit multiplication. This multiplier does not depend on the processor clock frequency, because both partial products and sums are calculated concurrently. The net advantage is that it eliminates the requirement of microprocessor to operate at increasingly higher clock frequencies. As the number of switching instances increases the operating frequency of a processor also increases. It leads to higher power consumption and also heat dissipation results in increased device operating temperature. The scalability T is the one more advantage of Vedic multiplier which is designed using Urdhva Tiryakbhyam sutra. As the input and output data bus widths increases the processing power also increases because of its regular structure. Due to its regular structure it can be easily layout in a silicon chip and consumes less area. If the number of input bits increases the gate delay increases drastically in conventional multiplier but it is slow in Vedic multiplier. Therefore the Vedic multiplier designed using Urdhva Tiryakbhyam sutra is efficient in terms of time, space and power.

The below figure explains the algorithm. Multiplication of 101 by 110.

1.First need to take right hand digits from both multiplicand and multiplier. Then multiply them together. Then we will get the LSB digit of the answer.

2. Multiply second bit of the top number with the LSB of the bottom number. Then multiply LSB of the top number with the second bit of the bottom number. Then add them together to get second bit of the answer.

3. Multiply third bit of the bottom with the LSB of the top number, second bit of the bottom number with the second bit of the top number, LSB of the bottom number with the third bit of the top number. Then add them together to get third bit of the answer.

4. This step is same as second step, just move one place to the left. We will multiply the second digit of one number by the MSB of the other number.

5. Finally multiply the LSB of the top and bottom number to get final product.

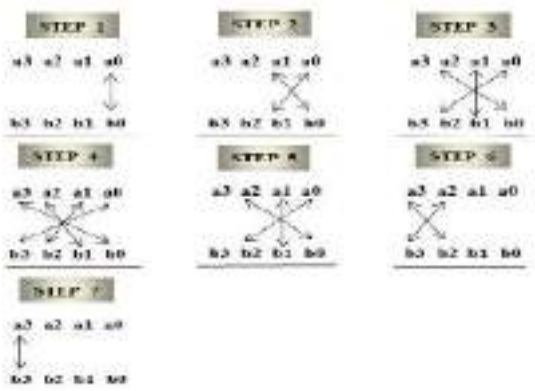


fig.4: multiplication of two 4 bit numbers using urdhvatiryakbhyam.

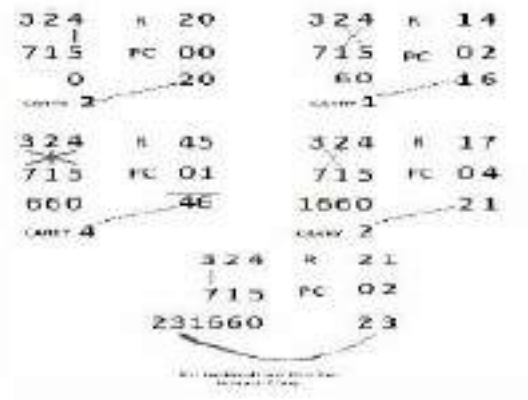


fig.5: urdhvatiryakbhyam algorithm for decimal multiplication.

III PROPOSED WORK

The hardware realization of 2*2 multiplier which is designed using conventional gates is shown in fig. the expression for the outputs i.e., q_0, q_1, q_2, $q_3$are shown in fig. The 2*2 multiplier which is designed using reversible gates based on Vedic mathematics is shown in fig. the circuit requires a total of 5 reversible logic gates out of five, 3 gates are Peres gates and remaining gates are Toffile gate and BME gate. The quantum cost of 2*2 Urdhva Tiryakbhyam multiplier is enumerated to be 23. The number of garbage outputs are 7. The number of constant inputs are 5.

The 4*4 reversible Vedic multiplier is designed using 2*2 Vedic multiplier. The block diagram of 4*4 Vedic multiplier is shown in fig. The block diagram consists of four 2*2 multiplier blocks each block has four bits as inputs. Two bits from the multiplicand and another two bits from multiplier. Two zeros are concatenated with the upper two bits and given as input to the four bit ripple carry adder are obtained from the second 2*2 multiplier. The output of the second and third 2*2 Vedic multiplier block is given to the 4-bit ripple carry adder as an input which is designed using reversible HNG gate. The output of second ripple carry adder is given to the first 4-bit ripple carry adder. The sum 2-bits from first four bit ripple carry adder is taken as output of the 4*4 Vedic multiplier. Another 2-bits sum output is given as an input to the third 4-bit ripple carry adder. The 4-bit output is taken from the fourth 2*2 Vedic multiplier and given as an input to the third 4-bit ripple carry adder. The output of this block indicates the output of last four bits of the 4*4 Vedic multiplier. In this design the ripple carry adder is designed using HNG gate. The quantum cost of this design is 166. The number of gates used in this design is 32. The number of constant inputs used in this design is 32. The garbage outputs in this design is 60.

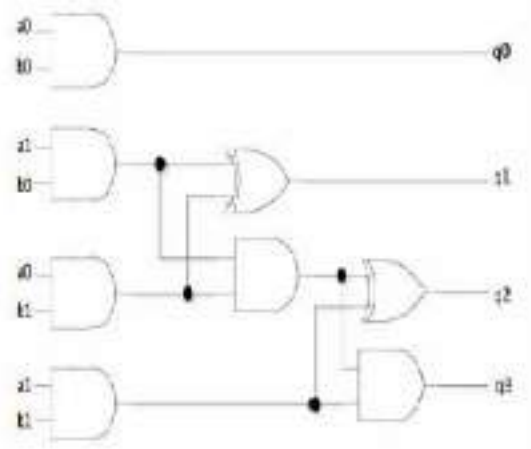


Fig. 6:2*2 Vedic multiplier using conventional gates.

q0= a0.b0
q1= (a1.b0) xor (a0.b1)
q2= (a0.a1.b0.b1) xor (a1.b1)
q3= a0.a1.b1.b0

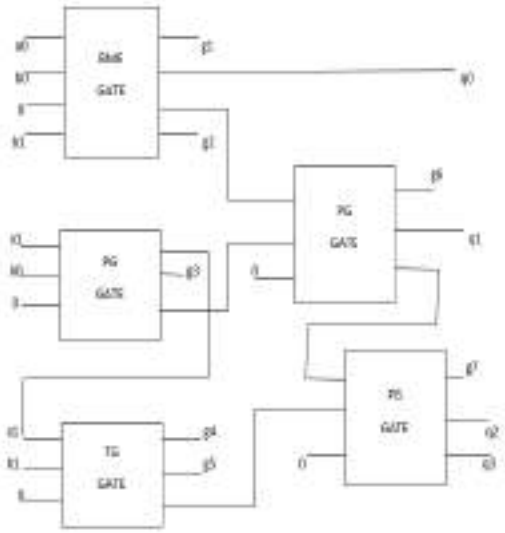


Fig.7: proposed 2*2 Vedic multiplier using Reversible logic gates.

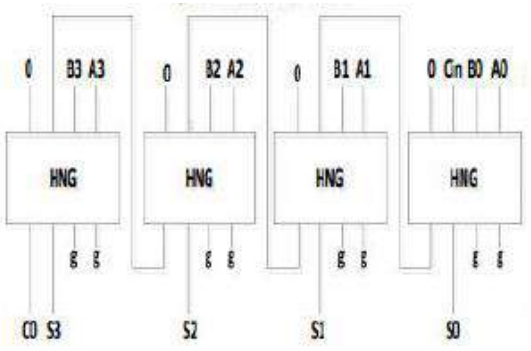


Fig.8: 4-bit Ripple Carry Adder units

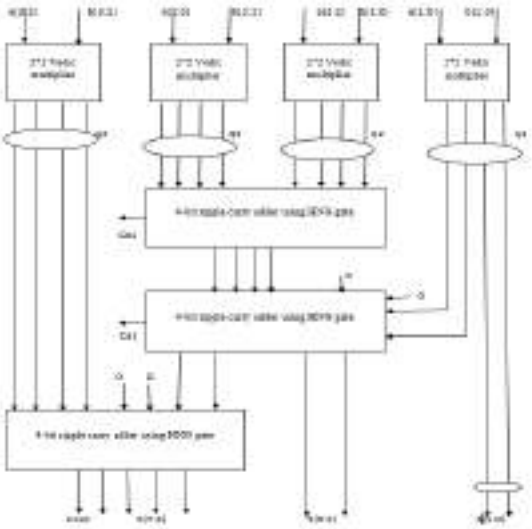


Fig.9: Block diagram of Reversible 4*4 Vedic multiplier.

IV RESULT ANALYSIS AND COMPARISION

The 2*2 and 4*4 reversible Vedic multiplier is designed and logically verified using XILINX 14.4 and Modelsim. The proposed 2*2 and 4*4 Vedic multiplier is implemented using VHDL coding.

Fig.10: Simulation results for reversible 2*2 Vedic multiplier.

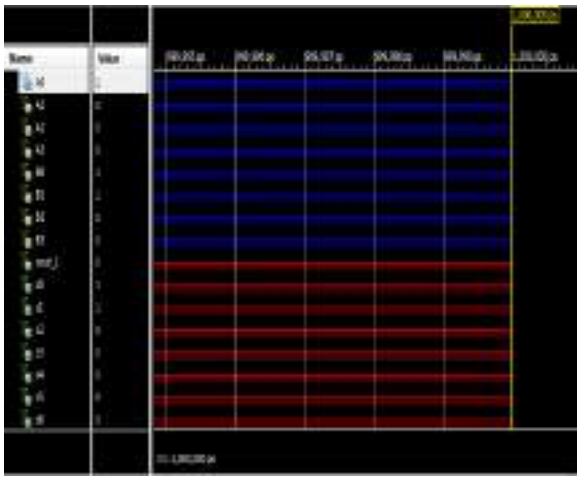


Fig.11: Simulation results for reversible 4*4 Vedic multiplier.

	Garbage output	Constant inputs	Gate counts	Quantum cost
Existing system 2*2	9	4	6	21
Existing system 4*4	62	29	37	162
Proposed method 2*2	7	5	5	23
Proposed method 4*4	60	32	32	166

Table.1: comparison of proposed method with the existing system.

V CONCLUSION

The proposed architecture of reversible Vedic multiplier exhibits speed improvements. The multiplier designed using Urdhva Tiryakbhyam sutra is efficient compared to existing system.

VI REFERENCES

[1] Design of High Speed low power Multiplier using Reversible Logic: Rakshith T.R.2013.

[2] G Ganesh Kumar and V Charishma, Design ofhigh speed vedic multiplier using Vedic mathematics techniques, ltn'l J. of Scientific and Research Publications, Vol. 2 Issue 3 March.

[3] M S Islam, M M Rahman, Z Begum and M Z Hafiz, 2009. Low Cost Quantum Realization of Reversible Multiplier Circuit. Information Technology Journal, vol. 8(2), pp. 208-213.

[4] Harpreet Singh Dhillon and AbhijitMithra "A Digital MultiplierArchitecture using Urdhva Tiryakbhyam Sutra of Vedic Mathematics" IEEE Conference Proceedings, 2008.

[5]d. Maslov, g. W. Dueck, and d. M. Miller, Synthesis of Fredkin-Toffoli Reversible Networks, IEEE Trans. VLSI Systems, **13(6)**, pp. 765769, 2005.

[6]P. D. Chidgupkar and M. T. Karad, "The Implementation of Vedic Algorithms in Digital Signal Processing", Global J, vol. 8, no.2, pp. 153-158, 2004.

[7]Monika Rangari et.al Design of reversible logic ALU using reversible logic gates with low delay profile international journal of advanced research in computer and communication engineering. Vol.4 issue 4. April 2015.

Design, Implementation and comparison of 8-Bit Vedic Multiplier Using Multiplexers and Logic Gates

Sumanashree C.V, M.C.Chandrashekhar, M Z Kurian
Dept. of ECE, Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India
sumanashree1234@gmail.com, mccshekhar@gmail.com, mzkurianvc@yahoo.com

ABSTRACT- In this paper, the design of high speed 8-bit vedic multiplier using multiplexers is presented. Vedic mathematic is the ancient Indian system of mathematics. It is mainly based on 16 sutras. In that 14th sutra is Urdhva Tiryakbhyam sutra which means vertical and crosswise. This sutra is used for designing of vedic multiplier in that the first design involves implementation of 8-bit Vedic multiplier using basic gates and the second design involves implementation of high speed 8-bit Vedic multiplier by using multiplexers. The system is designed using VHDL and it is implemented through Xilinx ISE 14.5.

Keywords— *Multiplier, Vedic multiplier, Vedic Mathematics, Urdhava Triyagbhyam*.

I. INTRODUCTION

Vedic mathematics is an ancient Indian system of mathematics consists of Vedic sutras .Vedic mathematics gives a simplified and optimized solutions compare to the conventional mathematical algorithms.

This Vedic algorithm are applied to a several mathematical operations [1] such as arithmetic, trigonometric, geometric operations in that one such operation is Vedic multiplier which is used for digital multiplier in digital signal processing applications.

Vedic mathematics consists of 16 sutras [2] which covers whole mathematical operations and digital multiplier based on Vedic multiplier are the fastest, reliable, efficient and low power multipliers by reducing the number partial products the delay also decreased and the system becomes faster.

From 16 sutras the 14th sutra is Urdhva Tiryakbhyam sutra Which is related for multiplication operation and this sutra is best suitable for lesser magnitude values.

II. URDHVA TIRYAKBHYAM SUTRA

In Urdhva Tiryakbhyam sutra operation are performed in vertical and crosswise manner. Multiplication operation is done by simple addition of partial products. In this sutra the parallelism architecture is used which means generation of partial products and their summing is performed simultaneously. As a result processor speed increases, calculations of Urdhva Tiryakbhyam sutra is shown in fig. 1.

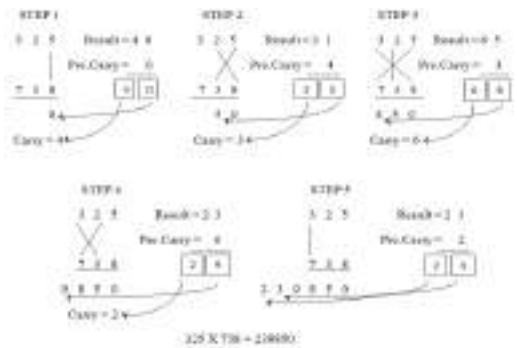


Fig.1: Multiplication of two decimal numbers by Urdhva Tiryakbhyam

Multiplication of two numbers 325x738 is performed by vertical and crosswise manner ,the result obtained by this multiplication is added with the carry generated from every pervious step and the procedure is continued till last bit of the input

III. 8-BIT VEDIC MULTIPLIER

The existing 8-bit vedic multiplier consists of four 4-bit vedic multipliers and three 8- bit ripple carry adder. Let us take A and B are of 8-bit input which gives an output S of sixteen bit and results are obtained after getting partial product and doing addition. Fig. 2 shows an existing 8-bit vedic multiplier.

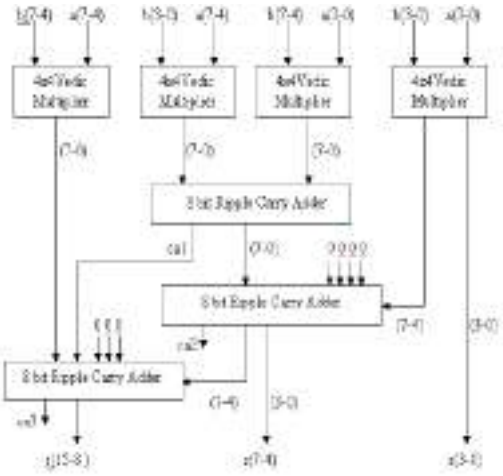


Fig.2: 8-bit Vedic multiplier

8-BIT VEDIC MULTIPLIER USING BASIC GATES:

The 8-bit vedic multiplier architecture is implemented by using logic gates, steps involved in implementation of vedic multiplier is first the design of 2-bit Vedic multiplier and second the design of 4-bit vedic multiplier for this four 2-bit vedic multiplier is used and they are interconnected by ripple carry adders. By using logic gates 2-bit vedic multiplier is shown in Fig.3

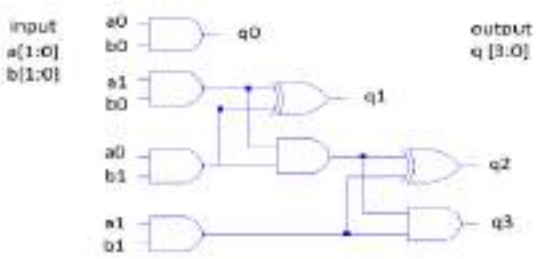


Fig.3: 2-bit Vedic multiplier using logic gates .

Let us take A and B are of 2-bit input which gives an output S of 4-bits , summing of partial products is performed by half adders same procedure is used in designing of 4-bit vedic multiplier in place of half adders ripple carry adders are used and is shown in Fig.2

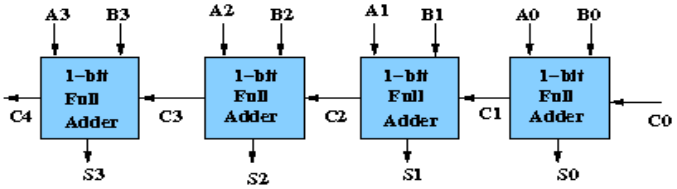


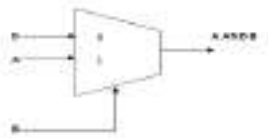
Fig.4: 4-bit ripple carry adder

The ripple carry adder consists of 2 inputs A and B of 4-bit and carry C of previous adder and produces an output S of 4-bits. By using two, four bit vedic multiplier and 8-bit ripple carry adder an 8-bit Vedic multiplier is designed as shown in Fig.4.

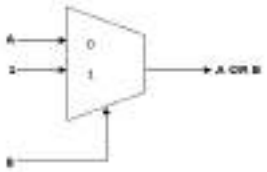
IV. 8-BIT VEDIC MULTIPLIER USING MULTIPLEXERS

The 8-bit vedic multiplier architecture is implemented by using multiplexers, all logic gates are replaced by multiplexers. Design of some of the basic gates by using multiplexers is shown.

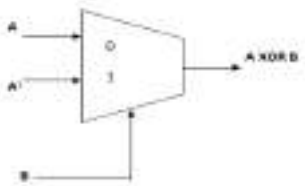
i. AND gate using multiplexer



ii. OR gate using multiplexer



iii. XOR gate using multiplexer



By using these multiplexers 2-bit vedic multiplier using multiplexers is designed by replacing AND, OR, and XOR gates by multiplexers is shown in Fig.5.

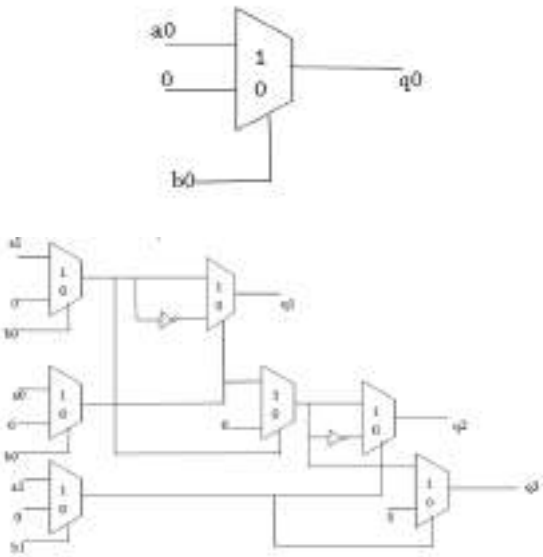


Fig.5: 2-bit Vedic multiplier using multiplexers.

Take 'a' and ' b' are of 2-bit input which gives an output 'q' of 4-bits here all the logic gates are replaced by multiplexers and summing of partial products is performed by half adders these adders also replaced by logic gates same procedure is used in designing of 4-bit Vedic multiplier in place of half adders ripple carry adders are used and they also replaced by multiplexers. By using two, four bit Vedic multiplier and 8-bit ripple carry adder using multiplexers an

8-bit Vedic multiplier using multiplexers is designed as shown in Fig.2.

V. SIMULATION RESULTS

The proposed 8-bit Vedic multiplier is coded in VHDL, simulated using Xilinx ISim simulator, synthesized using Xilinx XST and verified for the inputs a=00000011 and b=00000001which gives output of s=0000000000000011. The RTL schematic and simulation results of 8-bit Vedic multiplier are shown in Fig.6 and Fig.7.

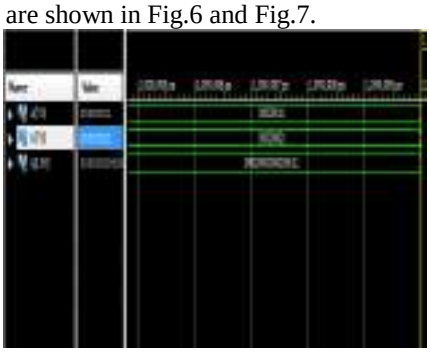


Fig.6: Simulation results of 8x8 Vedic multiplier using multiplexers.

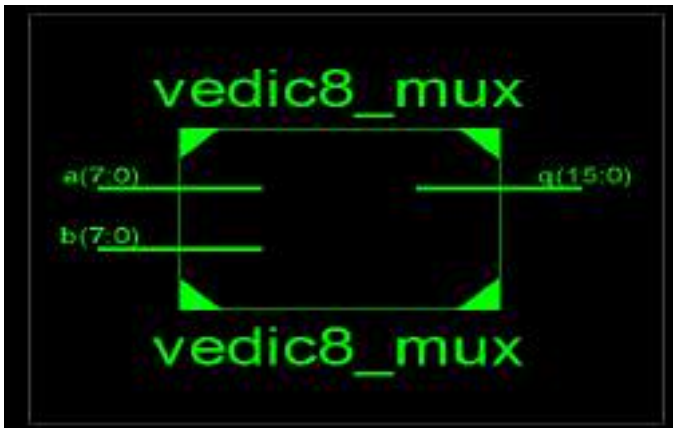


Fig.7: RTL diagram of 8x8 Vedic multiplier using multiplexers

Device utilization summary

```
Device utilization summary:
---------------------------

Selected Device : 6slx45csg324-3


Slice Logic Utilization:
 Number of Slice LUTs:                   26  out of  27288     0%
    Number used as Logic:                26  out of  27288     0%

Slice Logic Distribution:
 Number of LUT Flip Flop pairs used:     26
   Number with an unused Flip Flop:      26  out of     26   100%
   Number with an unused LUT:             0  out of     26     0%
   Number of fully used LUT-FF pairs:     0  out of     26     0%
   Number of unique control sets:         0

IO Utilization:
 Number of IOs:                          16
 Number of bonded IOBs:                  16  out of    218     7%
```

```
Total REAL time to Xst completion: 6.00 secs
Total CPU time to Xst completion: 6.12 secs

-->

Total memory usage is 186496 kilobytes
```

Timing Detail

```
All values displayed in nanoseconds (ns)

=========================================================================
Timing constraint: Default path analysis
  Total number of paths / destination ports: 336 / 8
-------------------------------------------------------------------------
Delay:               10.211ns (Levels of Logic = 7)
  Source:            a<0> (PAD)
  Destination:       q<6> (PAD)

  Data Path: a<0> to q<6>
                                        Gate     Net
    Cell:in->out      fanout   Delay   Delay  Logical Name (Net Name)
    ----------------------------------------  -----------
     IBUF:I->O             10    1.222   1.104  a_0_IBUF (a_0_IBUF)
     LUT4:I0->O             4    1.205   0.991  v2/m1/m1/Mmux_o111 (q2<1>)
     LUT6:I2->O             2    1.205   0.865  fa1/a3/m7/m1/Mmux_o111 (fa1/ca3)
     LUT5:I2->O             2    1.205   0.851  fa2/a4/m2/m1/Mmux_o111 (out2<3>)
     LUT6:I0->O             2    1.205   0.961  fa3/a1/m7/m1/Mmux_o111 (fa3/ca2)
     LUT5:I0->O             1    1.205   0.579  fa3/a3/m2/m1/Mmux_o111 (q_6_OBUF)
     OBUF:I->O                   2.571          q_6_OBUF (q<6>)
    ----------------------------------------
    Total                      10.211ns (5.820ns logic, 4.391ns route)
                                       (57.1% logic, 42.9% route)

=========================================================================
```

Table-1: Comparison between Vedic multipliers using basic gates and multiplexers.

Multiplier type	Vedic multiplier using basic gates	Vedic multiplier using multiplexers
Total Delay(ns)	17.102	10.211
Logic Delay(ns)	6.032	4.810
Route Delay(ns)	11.070	5.401
Total memory usage	196980 kilobytes	186496 kilobytes
Logic Levels	13	7

The8-bit vedic multiplier using multiplexers designed and compared with 8-bit vedic multiplier using basic gates in terms of total delay, logic delay ,route delay, total memory usage and number of logic levels .The results are tabulated in table 1 and it shows that there is a reduction in total delay, logic levels and memory usage. Thus, the proposed design is more efficient than the Vedic multiplier using basic gates.

VI. CONCLUSION

The 8-bitVedic multiplier architecture using multiplexers are designed by using Urdhva tiryagbhyam sutra. The total delay, memory usage and logic levels are reduced when compared with the design of vedic multiplier using basic gates.

REFERENCES

[1]. "Multiplier design based on ancient Indian Vedic mathematics" by Honey Durga Tiwari, Ganzorig Gankhuyag, Chan Mo Kim, Yong Beom Cho Dept. of Electronics Engineering Konkuk University Seoul, South Korea.

[2]. " Design and Implementation of High Speed 4x4 Vedic Multiplier" by a Debasish Subudhi Kanhu Charan Gauda, Abinash Kumar Pala, Jagamohan Das vol 4,Issue 11,novembee 2014 ISSN: 2277 128X.

[3] "Review P0aper on High Speed Karatsuba Multiplier and Vedic Mathematics Techniques" by Mr. Dharmendra Madke Assoc.Prof. Sameena Zafar Volume 3, Issue 12, December 2013 ISSN: 2277 128X

[4"]Simulation and Im1plementation of Vedic Multiplier using VHDL code" by G.Vaithiyanathan1,6, K.Venkatesan, S.Sivaramakrishnan, S.Siva and S. Jayakumar. Volume 4, Issue 1, January-2013ISSN 2229-55180.

Design Approach for Increased Lifetime of WSN using Artificial Neural Network Based Data Aggregation

Manomi K S, Vinutha C B, M Z Kurian
Department of ECE, SSIT, Tumkur, Karnataka
ksmanomi@gmail.com, cbvinutha@gmail.com, mzkurianvc@yahoo.com

ABSTRACT-In recent years, wireless sensor network (WSN) is finding its applications in agriculture, military, health care, environmental monitoring and industries. In wireless sensor networking energy resource of battery is very important and it has to be utilized in an efficient way. Energy resource of WSNs cannot be recharged or replaced in most of the time. So, the lifetime of the WSN is an important aspect in the design of WSN. Data aggregation is a crucial task in WSNs. This paper focuses on data aggregation using artificial neural network (ANN) to maximize the throughput of the network. The faulty sensor nodes in the network send the wrong data to cluster head (CH) by leaf nodes. This paper applies the neural network for wireless sensor network to detect the faulty sensor nodes and faulty nodes are eliminated. This increases the throughput and lifetime of sensor networks.

Keywords **-** Artificial Neural network (ANN), Cluster formation, Cluster head (CH), Data aggregation, Wireless sensor network (WSNs).

I. INTRODUCTION

Wireless sensor networks are composed of several sensor nodes. These sensor nodes senses the physical or environmental conditions like pressure, temperature, moisture, sound etc, and process the acquired data. Sensor nodes are connected wirelessly using sensor and actuators, passes the sensor node data through network to main location [1].

In WSN data aggregation is the process of accumulating and routing the information through a multi-hop network, processing the data at intermediate nodes to reduce energy consumption thereby increases the network lifetime.

Artificial Neural Network is an arithmetic algorithm to learn the complicated mapping between input and output. Data aggregation is the process of collecting and processing the sensor data in order to reduce the amount of data transmission in the network. Neural network based data aggregation increases the network lifetime and throughput.

A. Architecture of wireless sensor networks

Sensor nodes are placed in the sensor field as shown in the Fig 1. Each wireless sensor node has memory, communication device, controller, sensors/actuators and power supply [1]. These sensor nodes collect the data. The collected data is transferred to the base station through multi-hop path. Base station communicates with the user through the internet.

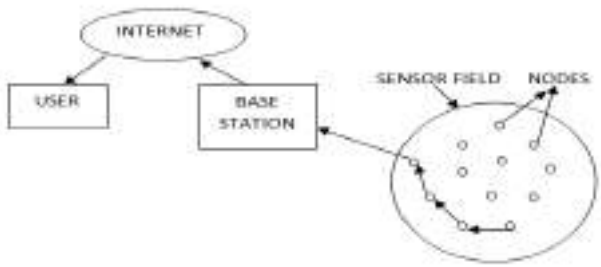


Fig. 1. Wireless sensor network architecture

II. ARTIFICIAL NEURAL NETWORK

The artificial neural network has processing units. These processing units are similar to biological neurons. Signals are transmitted through the connecting links [2]. Each links poses associated weights. Neural network may be single layer or multi layer network. Single layer ANN is shown in Fig.2. Here X_1, X_2 are inputs, W_1, W_2 are weights and Y is the output. The weights are multiplied with their incoming signals to get the output.

$$Y=X_1W_1+X_2W_2 \quad \text{..... (1)}$$

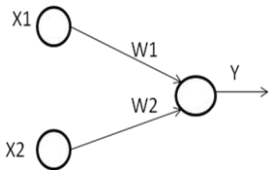


Fig. 2. Single layer ANN

Multi layer ANN has input layer, hidden layer and output layer as shown in Fig.3. Raw data is fed to the input neuron. Neurons in the hidden layer are found out by the activities of input neuron and associated weights between the input and hidden units. Neurons in the output layer depends on the activity of neurons in the hidden layer and connectiong weights between the hidden and output layer [2].

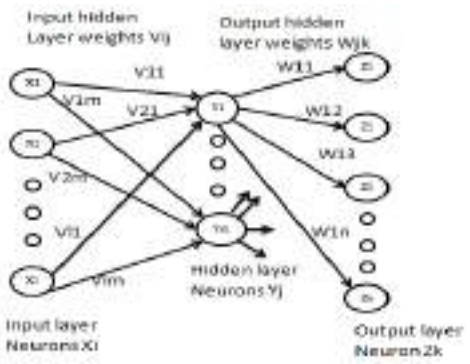


Fig. 3. Multi layer ANN

III. PROPOSED METHOD

A. Data aggregation

Data aggregation is a key technology in WSNs. Data aggregation is a process of dealing with the several data to obtain the data that must satisfy user needs. The sensor nodes are distributed in the specified sensor field. The sensor nodes are clustered using the clustering technique. We have mentioned the cluster heads for each cluster. The sensor nodes send their collected data from the environment to the cluster head and cluster head transfers the data to the base station using the data aggregation technique. In this paper ANN based data aggregation technique is used to increase the throughput and network lifetime.

The faulty sensor node data to the CH is the main drawback of the existing system [4]. The existing system is a sensor network without neural network. The data transfer in the existing system is shown in Fig.4. Fault tolerance is the ability of the sensor network to provide the correct results when the sensor nodes fail. The ANN for sensor network eliminates the faulty sensor node that avoids the false information processing.

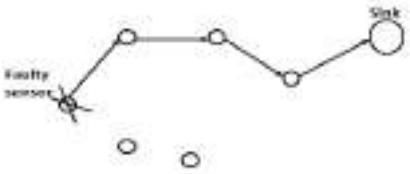


Fig. 4. WSN without artificial neural network

This paper presents an idea of the neural network for sensor network. The collected data from the sensors are input to the neural network. The neural network trains the sensed data and then output the neural network data. The neural network data are compared with the initial sensed data and the difference is α.

Artificial neural network data - Sensor data= α

α must be greater than β (β is the accuracy of production of data) otherwise the sensor is defective and that sensor is removed. The defected sensor node elimination is shown in Fig.5. This prevents the energy consumption for the transfer of false data.

Fig. 5. WSN artificial neural network

The failure of sensor node within the neural network does not cause the network failure because of the faulty sensor node elimination using artificial neural network. Therefore the network death doesn't happen with the faulty sensor.

B. Back propagation neural network for data aggregation

In this paper back propagation neural network (BPNN) training is used for the data aggregation. Back propagation neural network for data aggregation is as shown in Fig.6. BPNN is a multi layer feed forward neural network [2]. The input layer is located in the leaf nodes (cluster members) and hidden and output layer is in the cluster head (CH).This neural network involves supervised learning [8]. Back propagation network have:

1. Input layer consists of the back propagation network inputs.

2. Hidden layer consists of neurons; these neurons are responsible for adjusting the weights to determine the correct weights.

3. Output layer consists of back propagation network outputs and represents the final decision of training operation.

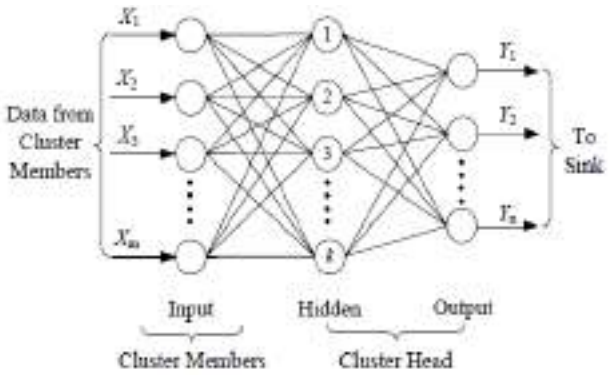


Fig. 6. Data aggregation using back propogation neural network.

In a feed forward neural network the information flows from leaf nodes to cluster head. Every node processes the data in following manner. The flowchart for data aggregation using BPNN is as shown in Fig.7. First initialize the weights for all nodes. Leaf node data are the input to the neural network. Then transfer the data to the cluster head for training or processing.

The aggregated data $X_j(n)$ of node C_j is given by

$$X_j(n) = y_j(n)w_j(n) + \sum_{K=k} Y_k(n)w_k(n) \quad \text{... (2)}$$

Where,

C_j = Cluster head

$w_j(n)$ =Associated weight of cluster head

$y_j(n)$ = Observation made by Cj

K = set of indices of all leaf nodes of node C_j.

$w_k(n)$ = Weights of leaf nodes

$Y_j(n)$= output after applying sigmoid function to $X_j(n)$

The sigmoid function is defined as $Y=1/(1+e^{-X})$ is used at intermediate nodes to get the binary decision. The decision $Y_j(n)$ of node C_j is given by,

$$Y_j(n)=1/(1+e^{-X_j(n)}) \quad \ldots\ldots (3)$$

The binary decision $Y_j(n)$ of each node is propagated to cluster head. The binary decision made by CH estimates the event hypothesis $H_{est}(n)=Y_{CH}(n)$.

The error at the CH is the error between the actual desired output H(n) and estimated output $H_{est}(n)$ is calculated as,

$$e(n)=H(n)-H_{est}(n) \quad \ldots\ldots..(4)$$

The error $e_j(n)$ at the node C_j is given by,

$$e_j(n)=e(n)w_j(n)Y_j(n)(1-Y_j(n)) \quad \ldots\ldots (5)$$

The gradient $grad_j$ is calculated to update the weight.

$$grad_j(n)=\alpha e_j(n)X_j(n)+\mu grad_j(n-1) \quad \ldots\ldots (6)$$

The value of step size α and acceleration factor μ are carefully selected to avoid algorithm divergence.

Finally the weights are updated as,

$$w_j(n+1)=w_j(n)+grad_j(n) \quad \ldots\ldots\ldots(7)$$

Use the updated weights for the next training sequence to train the algorithm which uses forward and back propagation till the convergence reaches [10].

If the error is more, then the node is defective. Eliminate the faulty node. When the leaf nodes fail or submit the faulty data, then reduce the nodes weight thereby resulting in the no impact of faulty data on aggregation. This reduces the energy consumption for the transfer of faulty data, thereby increases the network lifetime.

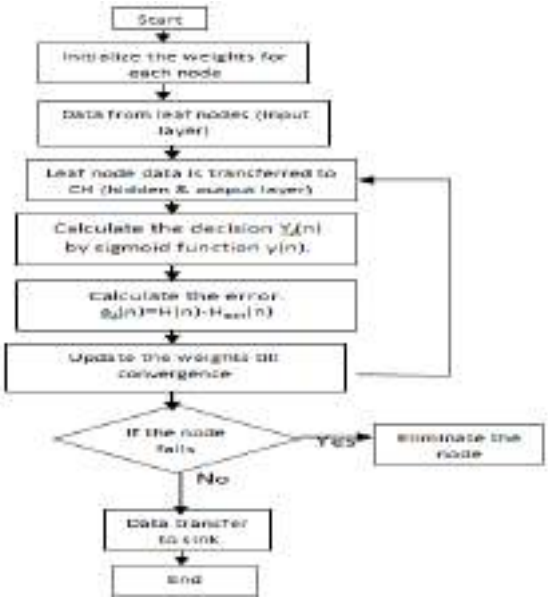


Fig. 7. Flowchart for data aggregation using back propogation neural network.

IV. RESULTS

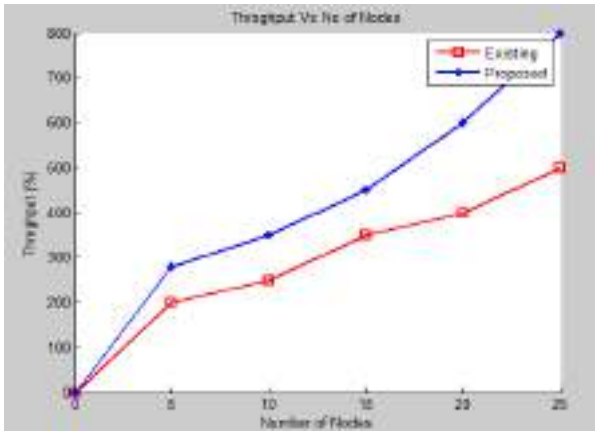


Fig. 8. Throughput Vs No of nodes

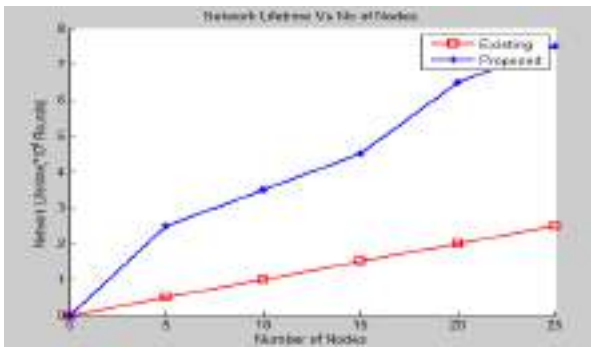


Fig. 9. Throughput Vs No of nodes

Due to the elimination of faulty sensor only real data is transferred. The throughput of the proposed system (with

neural network) is more compared to the existing system (without neural network) as shown in Fig.8. The network lifetime increase in a proposed system as shown in Fig.9. Comparison of the throughput and lifetime of existing system Vs proposed system is shown in the TABLE1.

TABLE 1: THROUGHPUT & LIFETIME COMPARISONS

NO OF NODES	WITHOUT NEURAL NETWORK		WITH NEURAL NETWORK	
	THROUGHPUT	LIFETIME *10^4 ROUNDS	THROUGHPUT	LIFETIME *10^4 rounds
05	200	0.5	300	2.5
10	240	1.0	350	3.5
15	330	1.5	440	4.5
20	400	2.0	600	6.5
25	500	2.5	800	7.5

V. CONCLUSION

Wireless Sensor Network is finding its application in various fields and is becoming very popular. Data communication in WSN should be supported by the limited resources of the nodes. Clustering of nodes in wireless sensor networks increases the scalability of the network and energy conservation. Data aggregation using ANN technique increases the network lifetime and throughput by eliminating the defected node. More efficient energy balancing technique should be devised to conserve nodes battery energy and result with improved network lifetime.

REFERENCES

[32] I. Akylidiz, W. Su, Sankarasubramaniam, and E.Cayrici, "A survey on sensor networks", IEEE Communications Magazine, Volume: 40 Issue: 8, August 2002, pp.102-114. Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. Received 12 December 2001; accepted 20 December 2001.

[33] Neda Enami, Reza Askari Moghadam, Kourosh Dadashtabar & Mojtaba Hoseini proposed "NEURAL NETWORK BASED ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORKS: A SURVEY". International Journal of Computer Science & Engineering Survey (IJCSES) Vol.1, No.1, August 2010.

[34] Shruti Prabha Shaktawat ,O.P Sharma, "Node Deployment Models and their Performance Parameters for Wireless Sensor Network: A Perspective" International Journal of Computer Applications (0975 – 8887) Volume 88 – No.9, February 2014.

[35] Krishanamachari B., D. Krishanamachari B., D. Estrin and S. Estrin and S. Wicker "The Impact of Data Aggregation in Wireless Sensor Networks", Distributed Event Based Systems (DEBS), Vienna, Austria, Jul. 2002

[36] Saeid Bahanfa, Helia Kousha[2] and Ladan Darougaran, "Neural networks for error detection and data aggregation in wireless sensor network", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011

[37] Vipin Pal, *Student Member, IEEE*, Girdhari Singh, and R. P. Yadav, "Balanced Cluster Size Solution to Extend Lifetime of Wireless Sensor Networks, IEEE INTERNET OF THINGS JOURNAL, VOL. 2, NO. 5, OCTOBER 2015

[38] Sang-Hyun Lee , Sang-Joon Lee , Kyung-Il Moon," Life-Environmental Sensor Data Aggregation Based on Neural Network", Advanced Science and Technology Letters Vol.47, ISSN: 2287-1233 ASTL Copyright © 2014 SERSC.

[39] Ling-Yi SuN, Wei CAI, Xian-Xiang HUANG "Data aggregation scheme using neural networks in w.ireless sensor networks", 978-1-4244-5824-0/$26.00 _c 2010 IEEE

[40] Jabal Raval, Bhushan Jagyasi "Distributed Detection in Neural Network based Multihop Wireless Sensor Network", 978-1-4799-2179-9/14/$31.00 ©2014 IEEE.

[41] Bin Zhang, Guolong Chen, "A Prediction-Based Fault-Tolerant Aggregation Algorithm in Wireless Sensor Networks" 2013 International Conference on Cloud Computing and Big Data.

Bhushan Jagyasi, Jabal Raval, "Data Aggregation in Multihop Wireless Mesh Sensor Neural Networks", 2015 ninth international conference on sensing technology.

ANTI-JAMMING TECHNOLOGY FOR MOBILE SIGNALS USING EMP JAMMER

Akash R Mannari
UG Scholar, Dept. of CSE, Rajarajeshwari College Of Engineering, Bengaluru, Karnataka, India
aakashmannari@gmail.com

ABSTRACT - 'Networking' being an important aspect of every individual's daily life process. We being in the era of Wireless Sensor Networks, the security will be considered as the major susceptible to numerous types of attacks. These attacks come under the category – Denial Of Service attack (DOS attack). Jamming is one such attack which is categorized under DOS attack. This processes at the physical layer, preventing the legitimate nodes which are performing its function i.e. producing signals, decreasing its network performance. Hence, Anti-Jamming techniques have become essential for ensuring proper delivery of measured event (network signal). Nowadays the wireless networks are being more affordable; the Anti-Jammer can be built to disrupt the operation of the Electromagnetic Pulse jammer (EMP Jammer -DOS attack).Resilience to electromagnetic jamming, EMP jammer, and their function are considered to be the difficult problem. It is often very hard to distinguish between the network breakdown and its challenge to conceal the activity pattern.
Basically in wirelesses networking system, needs are functioned to take common sleep duration to extend their battery life, resulting in well-organized patterns created by communication process. The patterns can be in predicted intervals which are jammed by EMP jammers. Hence, in this paper, an instrument is being introduced, named as 'Anti-Jammer'. For the sensor networks (EMP jammers) this mechanism is time-synchronized and modified in such a way that with higher frequencies the patterns get detached resulting in proper working of the jammed device. Through analysis, simulation and experimentation this paper demonstrates that the Anti-Jammer device's efficiency of any EMP jammer which has the lowest censorship-to-link utilization ratio.

Keywords- Anti-Jammer, EMP Jammer, Wireless networking, DOS attack, ARM7, RTC

I. INTRODUCTION

Communication jamming devices were initially developed and used for military purposes. Mobile phones' nowadays being a most necessary gadget for an every individual uses it everywhere. In some of the organizations like Indian Space Research Organization (ISRO), Nuclear Power Plant the usage of mobile phones is prohibited; hence to obsolete these they use "EMP JAMMERS", which is basically the electronic countermeasure device (EMC device) [1].

The technology used for cell phone jamming is very simple. The jamming device transmits an RF signal in the frequency range reserved for cell phones that interfere with the cell phone signal, which results in a "no network available" in your mobile display screen. The entire mobile phones network within that radius of the jammer is silenced. Using this technology any of the major organization can be signal-hacked (Gadgets using any kind of Radio wave signal can be hacked) very easily leading to a major chaos. Therefore, in this paper, the Antidote for the EMP JAMMER is being explained. This device jams the jammer and breaks down it. This method is called as "ANTI-JAMMER". Mentioning that cell phone jammers are illegal devices in most countries, this report is solely done for educational purposes.

II. RELATED WORKS

A. Mobile Phone Jammer

A mobile phone- network jammer is basically an instrument used to prevent cellular phones from receiving signals from base stations. When we use that, the jammer meritoriously incapacitates cellular phones. These devices can be used realistically in any location, but are found predominantly in places where a phone call would be particularly disruptive because of various reasons. As with another radio jamming, cell phone jammers block cell phone use by sending out radio waves along with the similar range of frequencies that a normal phone use. This causes adequacy interference with the communication between the mobile networks and towers to render the phones inoperable. On most retail phones, the network would simply emerge out of range. Most of the mobile phones utilize various other bands to transmit and obtain communications from a tower (called frequency division duplexing, FDD). Jammers can strive by either obstreperous phone to tower frequencies or tower to phone frequencies. Smaller handheld models block all bands from 800 MHz to 1900 MHz within a 30-foot to 40-foot range (9-10 meters). Small devices tend to use the former method while superior, more extravagant models may interfere directly with the tower [2] [3].

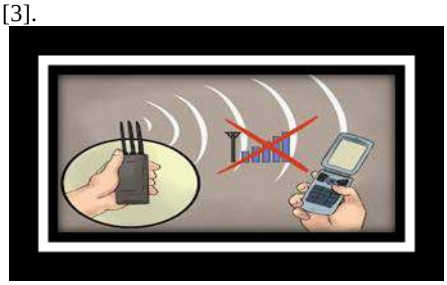


Figure1. Basic principle of Mobile jammer

The ambit of cell phone jammers can range from a dozen feet for mini models and micro models to kilometers for more extravagant models. The actual range of the jammer hinge on its power and the local surroundings, which may include any obstetrical that block the jamming signal. Less energy is required to disrupt signal from tower to a mobile phone than the signal from the mobile phone to the base station because the base station is located at larger distances from the jammer than the mobile phone and that is why the signal from the tower is not as strong. Earlier the EMP jammers were inadequate to working on mobiles phones which used only analog or any other older digital mobile phone standards. Newer models such as the double and triple band jammers can block all widely used systems (CDMA- Code Division Multiple Access, GSM- The Global System for Mobile Communication.) and are even very effective against newer phones which hop to different frequencies. [4][7].

Counter- Anti jammer encounters the jammed signal and breakdown the jammer circuit. The frequency will range more than 1800MHz resulting in breakage of the ionizer circuit used in the jammer, resulting in the normal functioning of the mobile networks.

1) Anti-Jamming Techniques:

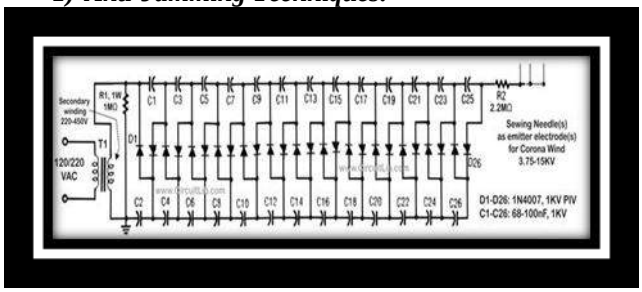


Figure2. Basic Circuit Diagram of an EMP JAMMER

When an anti-jammer is built the frequency generated will be extremely high, resulting in a breakdown of series connections of the capacitors. When the process of charging and discharging gets damaged the circuit of "EMP JAMMER" breaks down and hence acquires the original state and the mobile regains its original signals. There are several ways to counter jamming an RF device.

	UPLINK (Handset Transmit)	**DOWNLINK** (Handset Receive)
GSM 900	890-915 MHz	935-960 MHz
DCS 1800	1710-1785 MHz	1805-1880 MHz

The three most common techniques can be categorized as follows:

Table I: Operating frequency bands

Anti-Spoofing:

Basically in this kind of jamming, the EMP Jammer forces the mobile to switch off automatically by itself. This type is very difficult to be implemented since the EMP Jamming device first detects any mobile phone network in that specific surrounding area, then the device sends the signal to disable the mobile phone. Some of the types of similar technique can perceive if a nearby mobile phone is there and sends a message to tell the user to switch the phone to the silent mode. Through this, any person can control your mobile from anywhere.

Counter- Firstly the mobile gets switched off and then gets switched on again (for which the reason is unknown to the user too), by the time gadget gets restarted the operator will get the access to the users mobile. The Anti-spoofing process takes quite a time to target the mobile and process the anti-jamming technology with it.

Shielding attacks:

This is known as "TEMPEST" or "Electromagnetic Field shielding (EMF shielding)". This kind of attacks requires a closed circumference in a Faraday range so that any electronic signal generating device inside this range will not be able to transmit or receive RF signal from the outer surrounding of that range.

Counter- In this process, the EMF shielding will be eliminated through the device. Hence, the gadget gets recovered.

Denial of service:

This technique is referred to DOS. In this technique, the device transmits a noise signal at the similar operating frequency of the mobile phone network in order to drop off the signal-to-noise ratio (SNR) of the mobile within its minimal value. This method of EMP Jamming technique is soft head one since the device is always switched on [5]. This device is of this type; Mobile jammer circuit includes IF section, RF section, Antenna, and Power supply. GSM, used in digital cellular and PCS-based systems, operates at the 900-MHz and 1800-MHz bands in Asia. Jammers can broadcast on any frequency and are effective against Aviation Mission Planning System (AMPS), CDMA, Time Division Multiple Access (TDMA), GSM, Personal Communication Services (PCS), DCS systems [6][8].

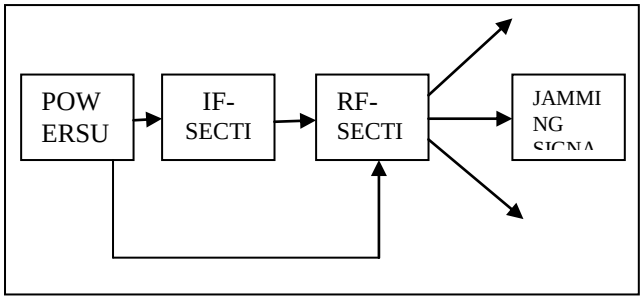


Figure3. Block diagram of JAMMER device

In our design, the jamming frequency must be the same as the downlink because it needs lower power to do jamming than the uplink range and there is no need to jam the base station itself. So, our frequency design will be as follows:

GSM 900 ----□ 960-1000 MHZ

GSM 1800---□ 1885-1960 MHZ

B. ARM 7(LPC2148)

The ARM7TDMI-S is a general purpose 32-bit microprocessor, which offers high performance and very low power consumption. The ARM architecture is based on Reduced Instruction Set Computer (RISC) principles, and the instruction set and related decode mechanism are much simpler than those of microprogrammed Complex Instruction Set Computers. This simplicity results in a high instruction throughput and impressive real-time interrupt response from a small and cost-effective processor core. Pipeline techniques are employed so that all parts of the processing and memory systems can operate continuously. Typically, while one instruction is being executed, its successor is being decoded, and a third instruction is being fetched from memory. The ARM7TDMI-S processor also employs a unique architectural strategy known as THUMB, which makes it ideally suited to high-volume applications with memory restrictions, or applications where code density is an issue. The key idea behind THUMB is that of a super-reduced instruction set. Essentially the ARM7TDMI-S processor has two instruction sets:

- The standard 32-bit ARM instruction set.
- A 16-bit THUMB instruction set.

The THUMB set's 16-bit instruction length allows it to approach twice the density of standard ARM code while retaining most of the ARM's performance advantage over a traditional 16-bit processor using 16-bit registers. This is possible because THUMB code operates on the same 32-bit register set as ARM code. THUMB code is able to provide up to 65% of the code size of an ARM, and 160% of the performance of an equivalent ARM processor connected to a 16-bit memory system.

C. REAL TIME CLOCK (RTC)

The real time clock (RTC) is a widely used device that provides accurate time and date for many applications. The RTC chip present in the PC provides time components of an hour, minute, and second in addition to the date/calendar components of a year, month, and day. The RTC chip uses an internal battery that keeps the time and date even when the power is off. One of the most widely used RTC chips is the DS1307 from Dallas semiconductor. The clock operates in either the 24-hour or 12-hour format with AM/PM indicator. The DS1307 has a built-in power-sense circuit that detects power failures and automatically switches to the backup supply. Timekeeping operation continues while the part operates from the backup supply.

D. Keypad

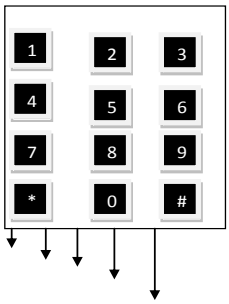


Figure4. 4x3 keypad structure

A basic 12 button keypad for user input. The buttons are set up in a matrix format. This allows a microcontroller to scan the 7 output pins to see which of the 12 buttons is being pressed. The jammer ON time and OFF time will be given with the help of keypad.

III. Result Analysis

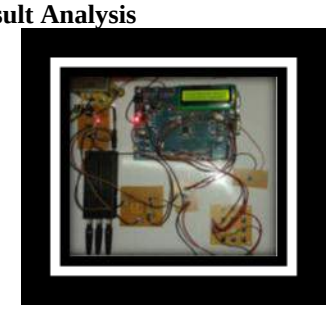


Figure5. Hardware design of the system

Figure6. Deactivation of mobile jammer and Activation of Anti-Jammer

Fig5 shows the hardware design of the mobile jammer with pre-scheduled time duration. It includes the Jammer, Anti-Jammer, ARM 7, keypad and relay. When the time schedule arrives in RTC the jammer will be activated with the help of relay and disrupt the communication system. This will be shown in fig6.

IV. CONCLUSION

This paper is successfully completed using Mobile jammer, Anti-Jammer, and ARM7. By this system, we can deactivate all the jammed signals at any location. This device can be applied in many places like;

- If the military Database is Jammed.
- In the War field, if the Radio signals are Jammed.
- In any of the organization where the signals might be jammed.

The design device works within the small range, with the proper installation and equipment; this can be transformed into a bigger device with a wider distance.

ACKNOWLEDGEMENT

The author gratefully acknowledges the support of management and Dr.Balakrishna R, Principal, RajaRajeshwari College of Engineering, Bengaluru, Dr.Usha Sakthivel, Professor, Head of Computer Science Department, RRCE, Bengaluru and my teachers and friends for their invaluable support and encouragement.

REFERENCES

[1] www.HowStuffWork.com

[2] En.wikipedia.org/wiki/Mobile_phone_jammer

[3] Multitopic conference2008.INMIC 2008.*IEEE* International

[4] "Zone of silence [cell phone jammer]," *Spectrum, IEEE, vol.42, no.5, 18*,May 2005

[5] Sami Azzam, Ahmad Hijazi, Ali Mahmoudy. "Smart Jammer for mobilephone systems"

[6] Mobile & Personal Communications Committee of the Radio Advisory Board of Canada, "Use of jammer and disabler Devices for blocking PCS,Cellular & Related Services"

[7]Ahmed Jisrawi, "GSM 900 Mobile Jammer", undergrad project, JUST,2006.

[8]John Scourias Overview of the global system for Mobile communications,http://ccnga.uwaterloo.ca/~jscouria/GSM/gsmreport.html#1

[9] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, (2002) "A survey on sensor networks", IEEE Communication. Mag., pp. 102-114.

[10]Cagalj M; Capkun S; Hubaux J.P; (2007) "Wormhole-Base Antijamming Techniques in Sensor Networks," Mobile Computing, IEEE Transactions on, vol.6, no.1, pp.100-114.

[11] P.Naresh, P. Raveendra Babu, K.Satyaswathi; Dept of ECE, CMR College of Engineering&Technology Hyderabad, AP-India. "International Journal of Science, Engineering and Technology Research (IJSETR)" Mobile Phone Signal Jammer for GSM, CDMA with Pre-scheduled Time Duration using ARM7.

Selective Harmonic Minimization method For Cascaded Multilevel Inverters using ANNs

Prakruthi Vasanth, Gowtham N
Department of Electrical and Electronics Engineering, VVCE, Mysuru-02
prakruthivasantha@gmail.com, gowtham.n@vvce.ac.in

ABSTRACT-The dc sources that feed the multilevel inverter are varying in time, and the switching angles should adapt to the dc source variations. Any variations in the input will result in harmonics at the output. To mitigate the harmonics in the inverter output, GA is used. A genetic algorithm optimization technique is applied to determine the switching angles for a cascaded multilevel inverter. Then, artificial neural networks are used to determine the switching angles for real time application. This eliminates specified order harmonics while maintaining the required fundamental output constant

Keywords - ANN, cascaded multilevel inverter, GA, neural network, real time, selective harmonic minimization, THD.

I. INTRODUCTION

MULTILEVEL inverters have drawn attention in recent years, especially in the distributed energy generation, because fuel cells, solar cells, wind turbines or micro turbines, batteries can be connected to feed the load or interconnect to the AC grid through multilevel inverters without any voltage fluctuations [1]. Multilevel inverters have a lower switching frequency and thus have reduced switching losses.

The output of multilevel inverter is in stepped form resulting in presence of harmonics. To reduce the harmonics different PWM schemes are suggested in the literature [2], [3]. But, PWM techniques increase the control complexity and the switching frequency. Another method to reduce the harmonics is by calculating the switching angles in order to eliminate certain order harmonics in the waveform [4]. Chiasson et al. [8] used the mathematical theory of resultants to calculate the optimum switching angles. But these expressions were high order polynomials that could not be solved when the number of levels is more.

In [10], analytical solutions for unequal dc sources have been obtained, and in [11], algorithms to solve for the angles have been derived. Particle swarm optimization is applied in [7] to reduce selective harmonics. All of these papers use intensive time-consuming equations to solve for the angles. Thus, the switching angle needs to be calculated offline.

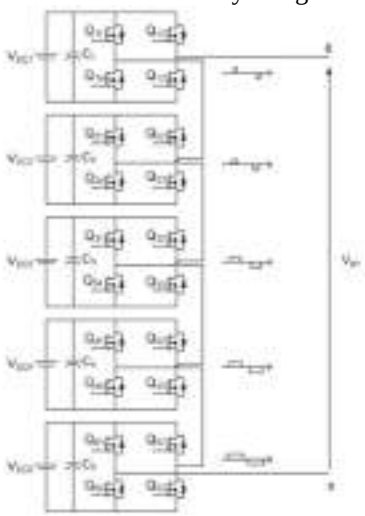
Liu et al. [13] have developed methods to calculate the switching angles in real time but, their approach was not extended for unequal dc sources.

An alternate approach to find the optimum switching angles in real time for varying dc sources is to calculate the switching angle offline using GA particle swarm optimization and storing the results in a lookup table. As the number of levels increases, the dimensionality and computational requirements increases. The solutions in the look up table can be used to train the ANN [14]. The target harmonics can also have an impact on the training of ANN. In this paper, the target harmonics for minimization are the 5th, 7th, 11th, and 13th.

This paper is organized as follows. In Section II, the multilevel inverters are explained. In Section III the problem is being described. In Section IV the set of equation for harmonic elimination is used to form the look up table using GAs. In Section V, the ANN-based control is explained, and in Section VI, experimental results are presented. Conclusions and final remarks are made in Section VII.

II. CASCADED MULTILEVEL INVERTERS

The cascaded inverter shown Fig. 1 (a) has many advantages such as circuit layout flexibility, additional stages can be easily added [5]. Here, five stages of H Bridge is used. The AC output comprises of stepped waveform with different switching angles as shown in Fig. 1 (b). The switching angles are considered to variables and should be determined by using GA.



1 (a)

Fig. 1. Cascaded inverter: (a) topology and (b) output waveform

III. PROBLEM DEFINTION

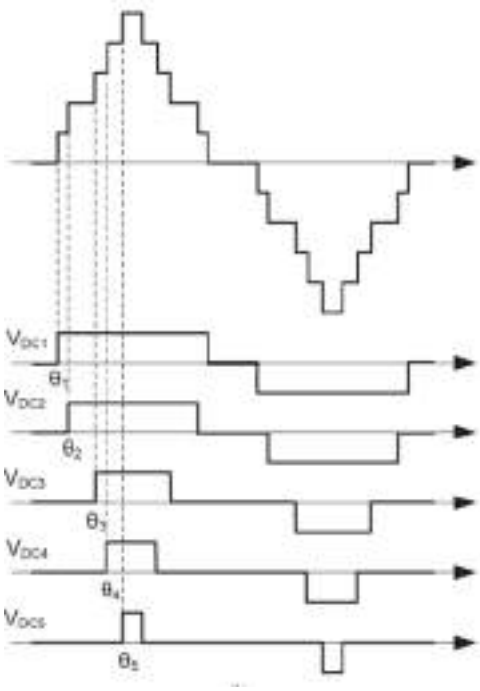
The demand for energy is increasing as a result there is increase in demand for Renewable Energy Sources (RES) such as solar cells, fuel cells and wind mills. Interfacing of RES in distributed generation and grid suffers voltage variations. This problem can be handled by adding converters [6]. Another approach to reduce voltage variations is by the use of switching frequency devices in inverters.

For example, the wind mill output will vary in accordance with the wind speed. Due to voltage fluctuations in wind speed, the fundamental output will not remain constant. The varying output results in harmonics. If the power converters alter the switching angles in accordance with wind speed, the output will be constant and harmonics are reduced.

The proposed methodology has two steps. First, the switching angles are calculated offline for predetermined input values through GA. Then, the GA is used to train the ANN for real time calculation of switching angles.

IV. GENETIC ALGORITHMS

The genetic algorithm (GA) is based on natural evolution and populations. GA is a population-based search method. The evaluation function represents a heuristic estimation of solution quality and the search process is driven by the variation and the selection operator. GA provides a general approach for searching for global minima or maxima within a bounded, quantized search space. Since GA only requires a way to evaluate the performance of its solution guesses without any a prior information, they can be applied generally to nearly any optimization problem [9]. The GA is composed of a fitness function, a selection technique, and crossover and mutation operators which are governed by fixed probabilities.



1 (b)

The optimum result is the end product containing the best elements of previous generations where the attributes of a stronger individual tend to be carried forward into the following generation. The GA iteration algorithm explores the search space using information from the points it has to bias the search towards the optimal point.

In a practical cases, in a multilevel inverter, all the dc sources vary to some degree. This variation can be due to the state of charge in a battery or fuel cell system, or it can be a function of solar irradiation in PV cells. In this way, it is necessary to monitor and control the switching angles to keep the desired output voltage characteristics. Under variations in the dc input sources, it is desired to maintain the fundamental output voltage. In applications for three phase machines, there is no need to cancel the harmonics that are a multiple of three because it is canceled in the line voltage. The set of equations for the GA is

$$V_{fund}=\frac{4}{\Pi}(V_{dc1}\cos(\theta_1)+V_{dc2}\cos(\theta_2) +\ldots+V_{dc5}\cos(\theta_5)) \quad (1)$$

$$V_{5th}=\frac{4}{\Pi.5}(V_{dc1}\cos(5\theta_1)+V_{dc2}\cos(5\theta_2) +\ldots+V_{dc5}\cos(5\theta_5)) \quad (2)$$

$$V_{7th}=\frac{4}{\Pi.7}(V_{dc1}\cos(7\theta_1)+V_{dc2}\cos(7\theta_2) +\ldots+V_{dc5}\cos(7\theta_5)) \quad (3)$$

$$V_{11th}=\frac{4}{\Pi.11}(V_{dc1}\cos(11\theta_1)+V_{dc2}\cos(11\theta_2) +\ldots+V_{dc5}\cos(11\theta_5)) \quad (4)$$

$$V_{13th}=\frac{4}{\Pi.13}(V_{dc1}\cos(13\theta_1)+V_{dc2}\cos(13\theta_2) +\ldots+V_{dc5}\cos(13\theta_5)) \quad (5)$$

The data set of angles to control the multilevel inverter for each different value of the dc sources using (1)-(5) was programmed with GA.

The equations (2)–(5) are set to be zero (no low-order harmonics). To solve (2)–(5) using GA, it is necessary to provide the real dc source values and the desired output voltage. The output voltage is set to 110 Vac, the data set is solved based on the different source value provided. After measuring the real values of the dc sources, a set of angles is found so that the output voltage is kept constant and the 5th, the 7th, the 11th, and the 13th harmonics are eliminated. A fitness function for the GA that evaluates each individual in the population was defined as follows:

$$f(V_{fund},V_{5th},V_{7th},V_{11th})=k_1|V_{fund}-110|+k_2|V_{5th}|+k_3|V_{7th}|+k_4|V_{11th}|+k_3|V_{13th}| \quad (6)$$

In (6), the coefficients k2 to k5 need to have higher value than k1, usually ten times greater for unbiased optimization. That procedure requires trial and error until a good balance is found. For this paper, k1 = 10 and k2 = k3 = k4 = k5 = 100.

The solution of the fitness function by GA requires 5input voltages to be varied for all its range. This is a trial and error problem since there is no previous information about the input voltages. For analysis the input dc voltage is varied between 24 to 40V.

V. ANN

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological

nervous systems, such as the brain, process information [16]. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements working in unison to solve specific problems. ANNs are generally time consuming to train but fast to run and can be easily parallelized once it is accordingly trained. Comparatively, lookup tables increase exponentially as the number of dc sources increases and it needs to deal with extrapolation leading to time-consuming algorithms and analytical approaches have to deal with the computational time required for the task.

The ANN topology proposed is shown in Fig. 2. It is a feedforward ANN with a tangent-sigmoid function activation hidden layer and a linear activation function output layer. This ANN takes the real dc source values and gives the switching angles for the control system.

The main reason for using ANN is to learn from the given data. Performance is evaluated during training as shown in Fig. 6. In this figure, for each topology, a number of training sessions are done and shown in this picture as the average value.

VI. SIMULATION RESULTS

In Fig. 3, the experimental results for an 11-level inverter operating with unequal dc sources are shown with the voltage values indicated. In Fig. 4, the frequency with a THD of 16.27%. In this same figure, a high value of the third and ninth harmonics can be noticed; those harmonics were not minimized due to the fact that they will be canceled in line voltage for a three-phase application.

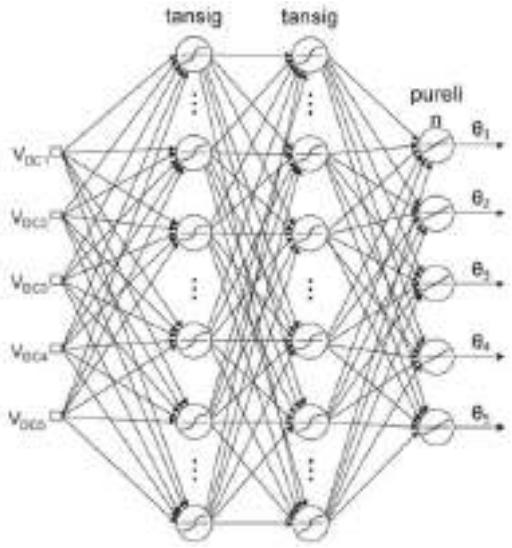


Fig. 2. Feedforward ANN topology

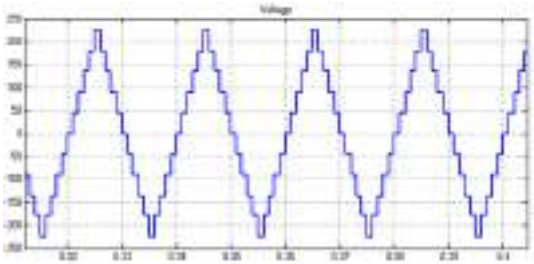


Fig.3. Multilevel inverter line output voltage waveform.

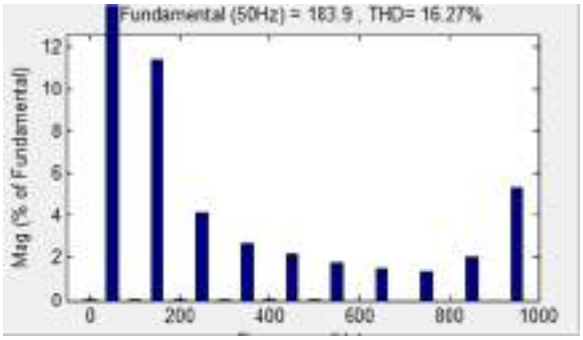


Fig.4. Output voltage frequency spectrum for Fig. 3

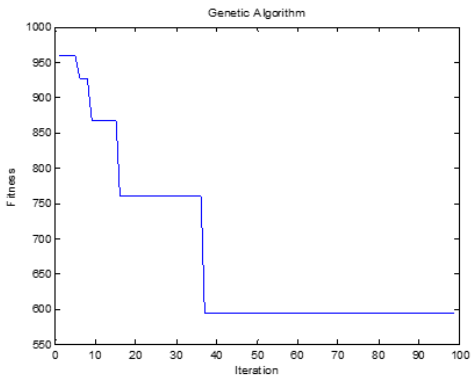


Fig.5. GA optimization

The output voltage as a function of dc source input variation and quality of solutions measured by GA's fitness value. The typical GA fitness curve for 100 iterations is shown in Fig. 5.The minimum value of fitness function is obtained after nearly 40 iterations. For the same values of dc sources, during real time operation the ANN takes only 4 iterations as shown in Fig. 6.

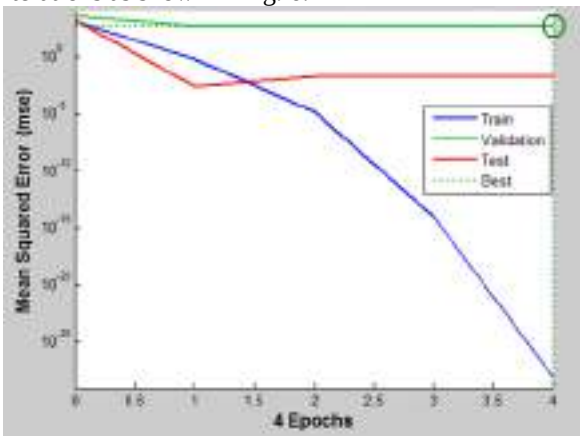


Fig.6. ANN performance results

The THD analysis of the output voltage after optimization is shown in Fig. 7. The THD obtained is less than 5%. Thereby improving the performance of the entire system.

Fig. 7. THD value after optimization

VII. CONCLUSION

A new method for real-time computation of switching angles using ANNs has been presented. The solutions were found offline using GAs to obtain the look up table for use during the training process of the neural network. The trained neural network is used then for online real-time determination of the angles. Experimental results were shown to validate this approach.

REFERENCES

[1] J. Chavarria, D. Biel, F. Guinjoan, C. Meza, and J. J. Negroni, "Energybalance control of PV cascaded multilevel grid-connected inverters under level-shifted and phase-shifted PWMs," *IEEE Trans. Ind. Electron.*, *vol. 60*, no. 1, pp. 98–111, Jan. 2013.

[2] M. S. A. Dahidah and V. G. Agelidis, "Selective harmonic elimination PWM control for cascaded multilevel voltage source converters: A generalized formula," *IEEE Trans. Power Electron.*, *vol. 23*, no. 4, pp. 1620–1630, Jul. 2008.

[3] Z. Du, L. M. Tolbert, and J. N. Chiasson, "Harmonic elimination for multilevel converter with programmed PWM method," in *Proc. IEEE Industry Applications Soc. Annu. Meeting, 2004*, pp. 2210–2215.

[4] T. Tang, J. Han, and X. Tan, "Selective harmonic elimination for a cascade multilevel inverter," in *Proc. IEEE Int. Symp. Ind. Electron*, Jul. 2006, vol. 2, pp. 977–981

[5] M. Malinowski, K. Gopakumar, J. Rodriquez, and M. A. P´erez, "A survey on cascaded multilevel inverters," *IEEE Trans. Ind. Electronics*, Vol. 57, No. 7, pp. 2197-2206, Jul. 2009.

[6] L.-S. Yang and T.-J. Liang, "Analysis and implementation of a novel bidirectional DC–DC converter," *IEEE Trans. Ind. Electron*., vol. 59, no. 1, pp. 422–434, Jan. 2012

[7] Baharuddin Ismail, Syed Idris Syed Hassan, Rizalafande Che Ismail, Abdul Rashid Haron and Azralmukmin Azmi, "Selective harmonic elimination of five-level cascaded inverter using Particle Swarm Optimization," *ISSN : 0975-4024 Vol 5* No 6 Dec 2013-Jan 2014**.**

[8] J. N. Chiasson, L. M. Tolbert, K. J. McKenzie, and Z. Du, "Elimination of harmonics in a multilevel converter using the theory of symmetric polynomials and resultants," *IEEE Trans. Control Syst. Technol.*, *vol. 13*, no. 2, pp. 216–223, Mar. 2005.

[9] D. E. Goldberg, Genetic Algorithms in Search, Optimization, Machine Learning. Reading, MA: Addison-Wesley, 1989.

[10] Z. Du, L. M. Tolbert, J. N. Chiasson, and H. Li, "Low switching frequency active harmonic elimination in multilevel converters with unequal DC voltages," in Conf. *Rec. IEEE IAS Annu. Meeting*, Oct. 2005, vol. 1, pp. 92–98.

[11] M. G. H. Aghdam, S. H. Fathi, and G. B. Gharehpetian, "Elimination of harmonics in a multi-level inverter with unequal DC sources using the homotopy algorithm," in Proc. *IEEE Int. Symp. Ind. Electron*, Jun. 2007, pp. 578–583

[12] Reza Salehi, Naeem Farokhnia, Mehrdad Abedi, and Seyed Hamid Fathi, "Elimination of low order harmonics in multilevel inverters using Genetic Algorithm*," Journal of Power Electronics, Vol. 11*, No. 2, March 2011

[13] Y. Liu, H. Hong, and A. Q. Huang, "Real-time calculation of switching angles minimizing THD for multilevel inverters with step modulation," *IEEE Trans. Ind. Electron., vol. 56*, no. 2, pp. 285–293, Feb. 2009.

[14] Faete Filho, Helder Zandonadi Maia, Tiago H. A. Mateus, Burak Ozpineci, Leon M. Tolbert, and João O. P. Pinto, "Adaptive selective harmonic minimization based on ANNs for cascade multilevel inverters with varying DC sources", *IEEE Trans. On industrial electronics, Vol. 60*, no. 5, May 2013

[15] K. El-Naggar and T. H. Abdelhamid, "Selective harmonic elimination of new family of multilevel inverters using genetic algorithms," *Energy Conversion and Management, Vol. 49*, No. 1, pp.89-95, Jan. 2008.

[16] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer, vol. 29*, no. 3, pp. 31–44, Mar. 1996.

DATA PROTECTION IN THE CLOUD BY USING JAR GENERATION

Hanusha.K , G.Priyanka

UG Student, SRM UNIVERSITY.

hanusha04@gmail.com , maahipriyanka111@gmil.com

ABSTRACT: Cloud computing enables highly scalable services to be easily consumed over the Internet on an as-needed basis. A major feature of the cloud services is that users' data are usually processed remotely in unknown machines that users do not own or operate. While enjoying the convenience brought by this new emerging technology, users' fears of losing control of their own data (particularly, financial and health data) can become a significant barrier to the wide adoption of cloud services. **Methods:** To address this problem, in this paper, we propose a novel highly decentralized information accountability framework to keep track of the actual usage of the users' data in the cloud. In particular, we propose an object-centered approach that enables enclosing our logging mechanism together with users' data and policies. **Findings:** We leverage the JAR programmable capabilities to both create a dynamic and traveling object, and to ensure that any access to users' data will trigger authentication and automated logging local to the JARs. To strengthen user's control, we also provide distributed auditing mechanisms. **Improvements:** We provide extensive experimental studies that demonstrate the efficiency and effectiveness of the proposed approaches.

Keywords: Cloud, JAR, CIA, LOGGER

1. INTRODUCTION

CLOUD computing presents a new way to supplement the current consumption and delivery model for IT services based on the Internet, by providing for dynamically scalable and often virtualized resources as a service over the Internet. Data handling can be outsourced by the direct cloud service provider (CSP) to other entities in the cloud and theses entities can also delegate the tasks to others, and so on1. Second, entities are allowed to join and leave the cloud in a flexible manner. As a result, data handling in the cloud goes through a complex and dynamic hierarchical service chain which does not exist in conventional environments2.

2. PROBLEM DEFINITION

2.1. Existing System

- Data handling in the cloud goes through a complex and dynamic hierarchical service chain
- This does not exist in conventional environments.
- Ordinary web framework
- Uses web services for request and responses3

2.1.2. Disadvantages

- No security for user's data. No authentication or security provided
- High resource costs needed for the implementation.
- Not suitable for small and medium level storage users4.

2. 2.Proposed System

- We propose a novel approach, namely Cloud Information Accountability (CIA) framework, based on the notion of information accountability.
- Our proposed CIA framework provides end-to end accountability in a highly distributed fashion
- It provides a detailed security analysis and discusses the reliability and strength of our architecture in the face of various nontrivial attacks by implementing Java Running Environment.
- In addition to a class file for authenticating the servers or the users, another class file finding the correct inner JAR, a third class file which checks the JVM's validity using oblivious hashing.
- Timer mechanism for limiting the accessing time for security purpose

- Secure JVM for making software tamper resistance capabilities to JAR file. It provides integrity, confidentiality to JAR.

2.2.1. Advantages

- One of the main innovative features of the CIA framework lies in its ability of maintaining lightweight and powerful accountability that combines aspects of access control, usage control and authentication.
- Providing defenses against man in middle attack, dictionary attack, Disassembling Attack, Compromised JVM Attack.
- It's Suitable for limited and large number of storages.

3. IMPLEMENTATION

3.1. JAR Generation

The JAR file contains a set of access control rules specifying whether and how the cloud servers and possibly other data interested party (users, companies) are authorized to access the content itself. Depending on the configuration settings defined at the time of creation, the JAR will provide usage control associated with logging, or will provide only logging functionality.

3.2. New user registration

In this module, the data owner uploads their data in the cloud server. The new users can register with the service provider and create a new account and so they can securely upload the files and store it. For the security purpose the data owner encrypts the data file and then store in the cloud. The Data owner can have capable of manipulating the encrypted data file. And the data owner can set the access privilege to the encrypted data file 5-7.

SCREENSHOTS

3. 3.Cloud service provider module

In Figure 1, The cloud service provider manages a cloud to provide data storage service. Data owners encrypt their data files and store them in the cloud with the jar file created for each file for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the cloud and then decrypt them

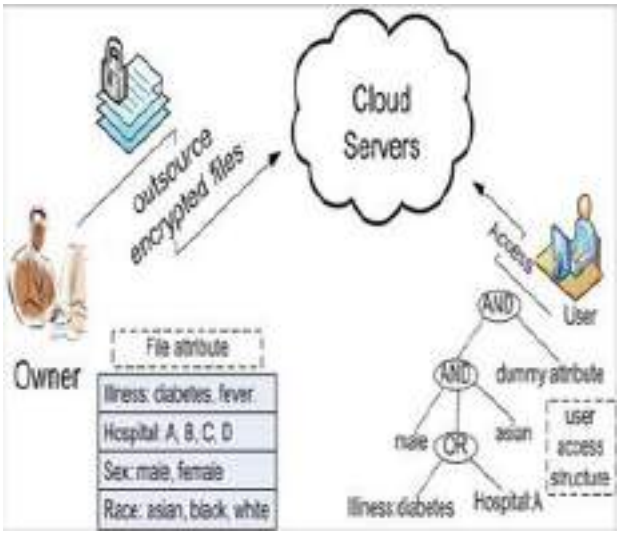


Fig 1. Cloud Server

3.4. Log Record Generation

We leverage the programmable capability of JARs to conduct automated logging. A logger component is a nested Java JAR file which stores a user's data items and corresponding log files. The main responsibility of the outer JAR is to handle authentication of entities which want to access the data stored in the JAR file.The data owner can specify the permissions in

user-centric terms as opposed to the usual code-centric security offered by Java, using Java Authentication and Authorization Services. Moreover, the outer JAR is also in charge of selecting the correct inner JAR according to the identity of the entity who requests the data is represented in Figure 2 .Log records are generated by the logger component. Logging occurs at any access to the data in the JAR, and new log entries are appended sequentially8.

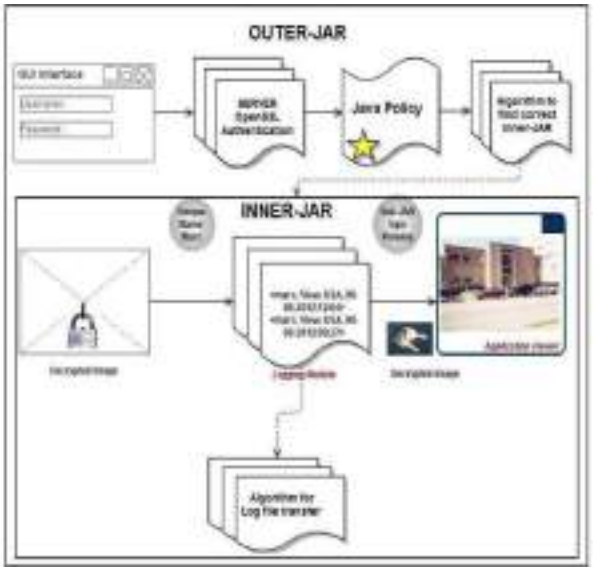


Fig 2. Components of JAR

SCREENSHOTS

3.5. Mode Setting

To allow users to be timely and accurately informed about their data usage, our distributed logging mechanism is complemented by an innovative auditing mechanism. We support two complementary auditing modes: 1) push mode; 2) pull mode.

Push mode. In this mode, the logs are periodically pushed to the data owner (or auditor) by the harmonizer. The push action will be triggered by either type of the following two events: one is that the time elapses for a certain period according to the temporal timer inserted as part of the JAR file; the other is that the JAR file exceeds the size stipulated by the content owner at the time of creation.

Pull mode. This mode allows auditors to retrieve the logs anytime when they want to check the recent access to their own data.

SCREENSHOTS

4. Technique in Proposed System

The Cloud Information Accountability (CIA) framework proposed in this work conducts automated logging and distributed auditing of relevant access performed by any entity, carried out at any point of

time at any cloud service provider. It has two major components: logger and log harmonizer.

There are two major components of the CIA, the first being the logger, and the second being the log harmonizer.The log harmonizer forms the central component which allows the user access to the log files. The logger is strongly coupled with users data.

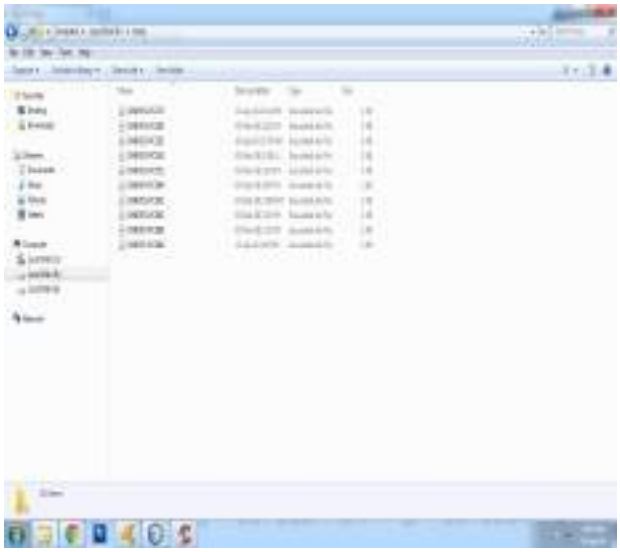
Figure 3.Data Flow Diagram

There are two major components of the CIA, the first being the logger, and the second being the log harmonizer.

The logger is the component which is strongly coupled with the user's data, so that it is downloaded when the data are accessed, and is copied whenever the data are copied. It handles a particular instance or copy of the user's data and is responsible for logging access to that instance or copy.

The log harmonizer forms the central component which allows the user access to the log files. The logger is strongly coupled with user's data (either single or multiple data items). Its main tasks include automatically logging access to data items that it contains, encrypting the log record using the public key of the content owner, and periodically sending them to the log harmonizer.

SCREENSHOTS

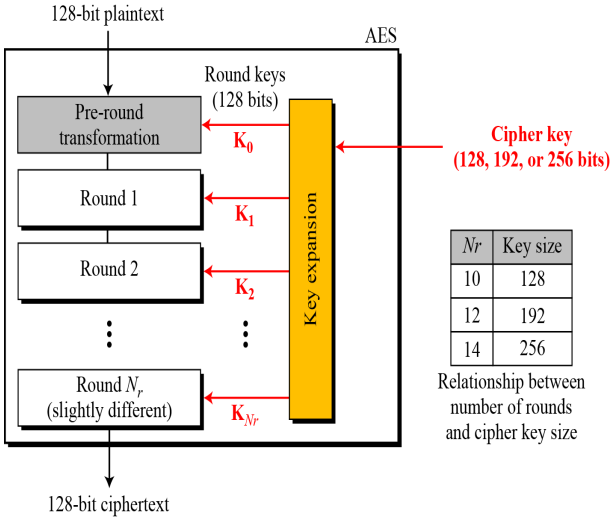


4.1.Algorithm in proposed system

Here we are using Advanced Encryption Standard to encrypt and decrypt the data. The Advanced Encryption Standard (AES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) in December 2001.

- The criteria defined by NIST for selecting AES fall into three areas:
- 1. Security
- 2. Cost
- 3. Implementation

- AES is a non-Feistel cipher that encrypts and decrypts a data block of 128 bits. It uses 10, 12, or 14 rounds. The key size, which can be 128, 192, or 256 bits, depends on the number of rounds.



Nr	Key size
10	128
12	192
14	256

Figure 4.AES Encryption

5.Conclusion

This system proposed innovative approaches for automatically logging any access to the data in the cloud together with an auditing mechanism. Our approach allows the data owner to not only audit his content but also enforce strong back-end protection if needed. Moreover, one of the main features of our work is that it enables the data owner to audit even those copies of its data that were made without his knowledge.

6.References

[1] P. Ammann and S. Jajodia, "Distributed Timestamp Generation in Planar Lattice Networks," ACM Trans. Computer Systems, vol. 11, 1993 Aug. pp. 205-225.

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc. ACM Conf. Computer and Comm. Security, 2007, pp. 598- 609.

[3] E. Barka and A. Lakas, "Integrating Usage Control with SIP-Based Communications," J. Computer Systems, Networks, and Comm., vol. 2008, pp. 1-8.

[4] D. Boneh and M.K. Franklin, "Identity-Based Encryption from the Weil Pairing," Proc. Int'l Cryptology Conf. Advances in Cryptology, 2001, pp. 213-229.

[5] R. Bose and J. Frew, "Lineage Retrieval for Scientific Data Processing: A Survey," ACM Computing Surveys, vol. 37, 2005 Mar. pp. 1- 28.

[6] P. Buneman, A. Chapman, and J. Cheney, "Provenance Management in Curated Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '06), 2006, pp. 539-550

[7].Yan Zhu,From RBAC to ABAC: Constructing lexible DataAccess Control for Cloud Storage Services, IEEE transactions on services computing, vol. 8, 2015 July, No 4

[8]Lan Zhou,"Trust Enhanced Cryptographic Role-Based Access Control for Secure Cloud Data Storage", IEEE transactions on information forensics and security, vol. 10, 2015 November, pp 11.

Enhanced AODV with Secrete Key Sharing to Improve Security and Energy Efficiency in MANETs

Mr. Murali G [#1], Ms. Divyashree B.N [#2], Dr. Balakrishna R [#3], [#4]Dr M.Vinayaka Murthy
[#1] *Research Scholar, Reva University,*
Associate Professor, Department of Information Science and Engineering
SJB Institute of Technology, 560060, Karnataka, India
[#2] *PG Scholar, CNE, Dept. of ISE*
SJB Institute of Technology, 560060, Karnataka, India
[#3]*Principal and Professor*
Raja Rajeshwari College of Engineering Bangalore -560074, Karnataka, India
[#4] *Professor*
School of Computer Science and Application, Reva University.

Abstract: Mobile Ad-hoc Networks (MANET) is one of the most important emerging research areas in the world. MANET nodes are miniature devices with limited energy, memory, transmission range, and computational power. Protecting the networks from different kinds of attacks i.e, security in wireless ad-hoc networks plays a decisive role and has received increased consideration in the current generation. As the goal of the MANET is to gather data from the deployed ad-hoc nodes and in-network processing, or data aggregation, the research.community are mainly concentrating in providing energy efficient paths in the network. Many protocols have been proposed in MANETs to provide security and energy efficiency. In this paper, by using Ant Colony Optimisation approach in AODV, the best energy efficient paths in the network are obtained. As nodes are portable in nature securing the network is also the main aim of the paper. The key distribution technique is used to aggregate and maintain the secrecy of data. The elliptic curve cryptography (ECC) data aggregation will enable the intermediate nodes for key distribution and key authentication.

Keywords— MANET, Data Aggregation, Elliptic Curve Cryptography, Ant Colony Optimization

I. INTRODUCTION

A MANET is a network consisting of numerous wireless nodes, which work together in determining some sort of physical or environmental conditions, such as temperature, sound, vibrations, light, movement etc. The individual ad-hoc nodes are small and have inadequate energy, computational power and memory which puts some fetters on the applications and protocols which are designed for use in such networks.

As these wireless ad-hoc nodes are tiny devices with restricted energy, reminiscence, transmission range, and computational power, a cluster head is usually used in the network, which receives the data from the nodes. Such a cluster head is usually a dominant computer with more computational power, energy and memory. These cluster heads can also be misinterpreted by the attackers. Therefore, some amount of security is required in order to maintain high survivability and reliability of the network.

Several nodes may be tasked with determining the phenomenon of collecting data, these nodes may cooperate in a "cluster" where one node is tasked with compressing the result from all the other nodes

in the cluster and produce a "collective view" of the cluster on the situation, which is called as data aggregation. Because of the nature of wireless ad-hoc networks, all nodes in the network may not have a direct association with other nodes in the network. Hence, they make use of multi-hop communication in order to communicate. Multi-hop communication in wireless ad-hoc networks is expected to consume less power than the traditional single hop communication, which is also enviable in order to keep the communication costs at its minimum.

Ad-hocs are often deployed in easily reached areas, which add the risk of physical attack. This is why wireless ad-hoc networks pose inimitable challenges. Attacks can be internal or external kind. The attacker can be either active or passive. There are different attacker types which are considered in the paper is black hole attack. The black hole is a node which acts as a normal node by advertising the path to destination but discards all the packets it receives.

By minimizing delay in packets delivery, the life time of the network and Quality of Service of the network can be improved, which can also be extended by using suitable energy efficient routing protocol. The design of a Wireless Ad-hoc Network depends on the application, and it must deliberate factors such as the environment, the application's design objectives, and system constraints.

II. RELATED WORKS

Several energy aware routing protocols are designed based ACO heuristics. Gupta et al[1] did a comparison of three ACO-based protocols : Ant-AODV, Ant-DSR and Ant-DYMO based on different performance metrics like routing overhead, end-to-end delay and few other many. Radwan et al [2] proposed AntNet RLSR, in which mobile agents build routes between source and destination simultaneously discovering network activities and updating routing table. Zhu et al. [3], for asymmetric schemes focused on preserving data integrity and proposed an

efficient integrity-preserving data aggregation protocol named EIPDAP. The scheme is based on the modulo addition operation using ECC, and has the most optimal upper bound on solving the integrity-preserving problem for data aggregation. Niu et al. [4] proposed a secure identity-based lossy data aggregation scheme using homomorphic hashing and identity-based aggregate signature. In the scheme, the authenticity of aggregated data can be verified by both aggregators and BS. The computation and communication overheads could be significantly reduced because the BS can perform batch verification. However, the above two schemes may lead to the leakage of data privacy due to decryption at the aggregator.

West off et al. [5] Based on PH, proposed CDA methods to facilitate aggregation in encrypted data, where richer algebraic operations can be directly executed on encrypted data by aggregators.

Mykletun et al. [6] adopted several public-key-based PH encryptions to achieve data concealment in WSNs. Furthermore. Girao et al. [7] proposed a novel scheme by extending the EL Gamal PH encryption.

III. Proposed System

In this section, Architecture of the proposed system is explained. It includes Network Initialization, Cluster formation, Key distribution, E-AODV routing, Data Aggregation and Data Authentication.

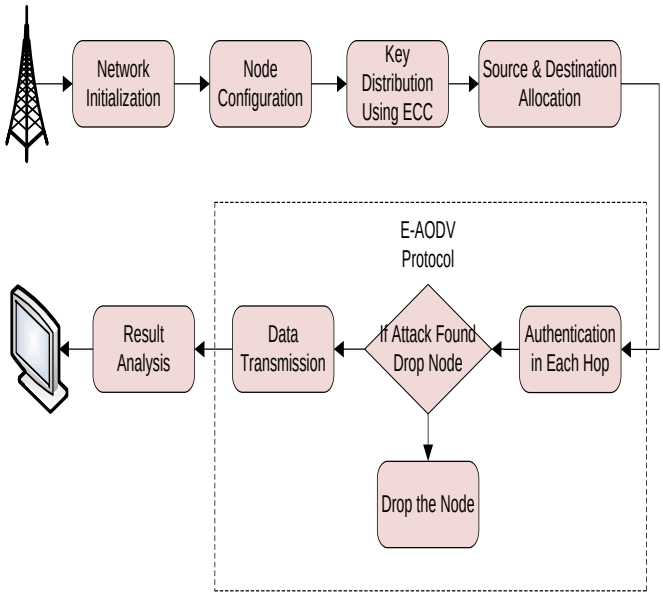


Figure 1.Block Diagram of Proposed System.

a. *Data Aggregation*

Data aggregation is one of the techniques to effectively utilize the limited resources. Generally it consists of the following steps :(i) Cluster head selection(ii) Cluster group formation. (iii) Transfer of data. There are few protocols that can perform the data aggregation and routing simultaneously. In each cluster, cluster head is chosen to collect the data from all the ad-hoc nodes and aggregate and send it to the requested base station depending on the energy of each node. Here we make use of a LEACH-KED[3] algorithm to form the clusters. Once clusters are formed, data aggregation process will begin. Then the data will be transmitted to the destination securely using ant colony optimised demand routing protocol Ad-hoc on-Demand Distance Vector (E-AODV) for a secure hop-by-hop data aggregation. In this, data is encrypted by the transferring nodes and decrypted by the aggregator nodes which then aggregate the data and encrypt the aggregation result again. At last the sink node gets the final encrypted aggregation data and decrypts it. Aggregate nodes are vulnerable to attack because of decryption of the ad-hoc data's in it.

b. *Ant Colony Optimization in AODV*

Ant Colony Optimization (ACO) is the foraging behaviour of real ants that tries to find the shortest paths between food sources and the nest which is helpfull in finding the optimised paths to the destination. The ACO's amalgamation is also motivated by the usual good performance shown by the algorithms using it. The ACO in AODV works as follows:

1. The route discovery phase is initiated when the source node initiates to send message across the network by broadcasting the route request packet which are forwarded to the neighbouring nodes in the network. These neighbour nodes forward the request packet to their nodes till the destination node is reached.
2. Nodes while forwarding the route request packet checks for the residual energy of the neighbouring nodes and select for the routing path in the network.
3. As the nodes lose energy, those nodes will be dropped and the alternative nodes with high residual energy will be selected for routing.
4. After the route request packet reaches the destination, authentication part is done and the route reply packet is generated and traversed back to the source node.
5. After the source node receives the route reply packet from the destination, the source node records the path in the table. If the source node receives multiple route replies, the route with optimized path is selected for data transmission.

If any link failure occurs in the network during data transmission, the route error packet is generated and transmitted source node and an alternative path for

destination is done. The energy of the nodes are determined by the below formula

If((Node N< 0.9* E_avg)

then(drop RREQ)) , (1)

The nodes that satisfies the above condition are considered during data transmission. The nodes that do not have sufficient energy are discarded until and unless they possess the required amount of residual energy for the capability of transmission.

c. Data Authentication

It is the confirmation of sender or receiver. It ensures that communicating node is the one that it has to communicate with.

In this paper, using Elliptic Curve Cryptography (ECC) which is an approach to public-key cryptography that is based on the algebraic structure of elliptic curves over fixed fields. Public-key cryptography is based on the intractability of firm mathematical problems. Earlier public key systems, such as RSA algorithm, are secure assuming that its difficult to factor a large integer composed of two or more large prime factors. For elliptic-curve-based protocols, it is believed that finding the discrete logarithm of a random elliptic curve element with respect to a publicly-known base point is infeasible. The size of the elliptic curve determines the complexity of the problem. It is believed that, the level of security afforded by an RSA-based system with a large modulus can be achieved with a much smaller elliptic curve group by using a small group that reduces storage and transmission requirements.

SP-ECC Message Authentication: In this section, we propose an absolutely secure and efficient SP-ECC. The main idea is that for each message m to be unconfined, the message sender, or the sending node, generates a source unnamed message authenticator for the message m. The generation is based on the MES scheme on elliptic curves. For a ring signature, each ring member is required to compute a forgery signature for all other members in the AS. In our scheme, the entire SP Massage authentication generation requires only three steps, which link all non-senders and the message sender to the SP alike. In addition, our design enables the SP-ECC message authentication to be verified through a single equation without individually verifying the signatures.

Let p > 3 be an odd prime. An elliptic curve E is defined by an equation of the form:

$$E: y^2 = x^3 + ax + b mod p, \quad (2)$$

Where $a, b \in F_p$, and $4a^3 + 27b^2 \neq 0 mod p$. The set $E(F_p)$ consists of all points $(x, y) \in F_p$ on the curve, together with a special point O, called the point of infinity.

Let $G = (x_G, y_G)$ be a base point on $E(F_p)$ whose order is a very large value N. User A selects a random integer $d_A \in [1, N-1]$ as his private key. Them, he can complete his public key Q_A from $Q_A = d_A \times G$.

Signature generation algorithm as follows.

Step1: select a random integer k_A, $1 \leq k_A \leq N-1$.

Step2: Calculate $r = x_A mod N$, where $(x_A, y_A) = k_A G$. If r=0, go back to step 1.

Step3: Calculate . If r=0, go back to step 1.

Step3: Calculate $h_A \underset{\leftarrow}{l} h(m, r)$, where h is a cryptographic has function, such as SHA-1, and $\underset{\leftarrow}{l}$ denotes the l leftmost bits of the hash.

Step4: Estimate $s = r d_A h_A + k_A mod N$. If s=0, go back to step 2.

Step5: the signature is the pair (r, s).

Signature Verification algorithm is as follows:

Lets say for Sender authenticate the receiver signature, the receiver must have the public key Q_A, then the algorithm steps follows as shown below.

Step1: Checks that $Q_A \neq 0$, otherwise invalid

Step2: Checks that Q_A lies on the curve.

Step3: Checks that $nQ_A = 0$

d. Attack Model

Black hole attack this is a DoS attack, where a malicious node advertises a zero cost route through itself. If the routing protocol in the network is a "low cost route first" protocol, like distance vector, other nodes will chose this node as an intermediate node in routing paths. The neighbours of this node will also chose this node in routes, and compete for the bandwidth. This way the malicious node creates a black hole inside the networks.

IV. RESULTS

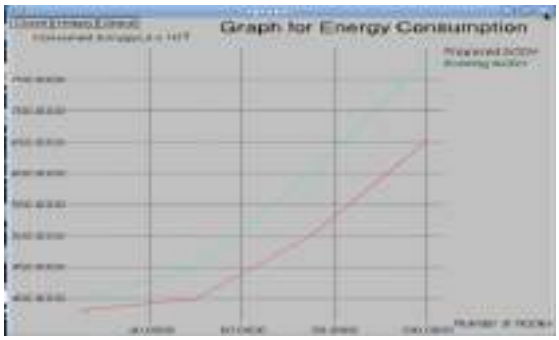


Figure 2: Energy Consumption as the number of nodes increases.

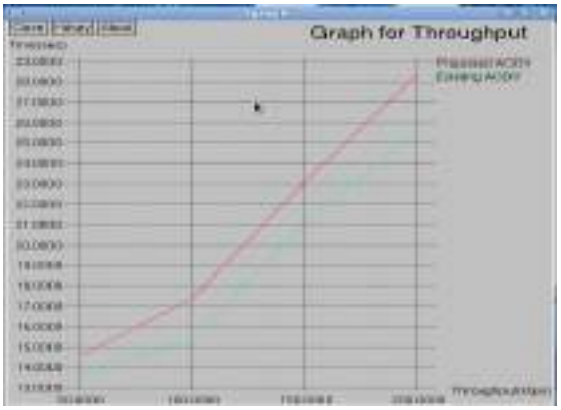


Figure 3: Throughput

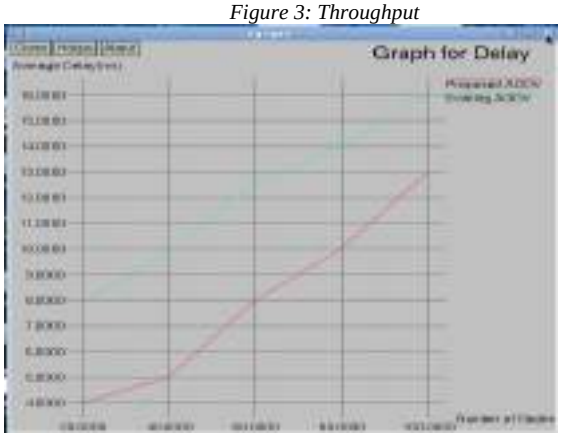


Figure 4: Delay.

Figure 5: Packet Delivery Ratio.

V. CONCLUSION

The main Objective of the proposed system is to increase the network lifetime by reducing the energy consumption of ad-hoc nodes during data transmission. In this paper, ant colony optimization with AODV as well as encrypted-data aggregation scheme based on elliptic curve cryptography that exploits a smaller key size are proposed. Additionally, it allows the use of higher number of operations on cipher-texts and prevents the distinction between two identical texts from their cryptograms. These properties permit the approach to achieve higher security levels than existing cryptosystems in ad-hoc networks. Performance evaluation shows the proposed system works much better than existing system in terms of energy consumption, throughput, delay and packet delivery ratio.

References

[1] A. K. Gupta, H. Sadawarti and A. K. Verma, "MANET Routing Protocols Based on Ant Colony Optimization", Intl. Journal of Modelling and Optimisation, vol 2, pp. 42-49, feb 2012.

[2] A. A. Radwan, T. M. Mahmoud and E. H. Hussein, "AntNet-RSLR: A proposed Ant Routing protocol for Manets", Proc, of IEEE Saudi Intl. Electronics, communication and Photonics Conference Riyadh, Saudi Arabia, pp 1-6, Apr 2011.

[3] Zhu et.al[1], " An Efficient data aggregation protocol concentrated on data integrity in wireless ad-hoc networks." Int. J. Distrib. Sens. Netw. 2013, 2013, 256852.

[4] Niu, et.al[2] "Lossy data aggregation integrity scheme in wireless ad-hoc networks." Comput. Electr. Eng. 2013, 39, 1726–1735.

[5] Westhoff,et.al[3]," Concealed data aggregation for reverse multicast traffic in ad-hoc networks: Encryption, key distribution, and routing adaptation". IEEE Trans. Mob. Comput. 2006, 5, 1417–1431.

[6] Mykletun, E.; Girao, J.; Westhoff, D. Public key based cryptoschemes for data concealment in wireless ad-hoc networks. In Proceedings of the IEEE International Conference on Communications (IEEE ICC '06), Istanbul, Turkey, 11–15 June 2006; pp. 2288–2295.

[7] F. Ye, H. Lou, S. Lu, and L. Zhang, "Statistical En-Route Filtering of Injected False Data in Ad-hoc Networks," Proc. IEEE INFOCOM, Mar. 2004.

[8] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An Interleaved Hop-By-Hop Authentication Scheme for Filtering False Data in Ad-hoc Networks," Proc. IEEE Symp. Security and Privacy, 2004.

[9] M. Albrecht, C. Gentry, S. Halevi, and J. Katz, "Attacking Cryptographic Schemes Based on 'Perturbation Polynomials'," 2009.

[10] Lejla Batina, Nele Mentens, Kazuo Sakiyama, Bart Preneel, and Ingrid Verbauwhede, "Low-Cost Elliptic Curve Cryptography for Wireless Ad-hoc Networks", Springer, pp. 6–17, 2006.

[11] Sweta Nigam and K.N. Hande, "Survey on "Security Architecture Based on ECC (Elliptic Curve Cryptography) in Network", International Journal of Computer Science and Mobile Applications, Volume 3, Issue 1, 2015.

[12] Mehala.G and Mariselvi.J, "Source Anonymous Message Authentication in Security Networks", International Journal of Computer Trends and Technology, volume 17, Issue 3, 2014.

[13] Chinnaswamy C.N and Natesha B V, "Message Authentication between the Nodes using modified El-Gamal Signature on Elliptic Curve", International Journal for Advance Research in Engineering and Technology, Volume 2, Issue V, 2014.

[14] Naipunya H C, Nalina G R, Gururaj H L and Ramesh B, "Secured Source Anonymous Message Authentication Using Wireless Ad-hoc Network", Journal of Computer Engineering, Volume 17, Issue 3, 2015.

[15] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on ad-hoc networks, IEEE Commun. Mag. 40 (8) (2002) 102–114. [2] J. Yick, B. Mukherjee, D. Ghosal, Wireless ad-hoc network survey,Comput. Networks 52 (12) (2008) 2292–2330. [3] K. Akkaya, M. Demirbas, R.S. Aygun, The Impact of Data Aggregation on the Performance of Wireless Ad-hoc

Networks, Wiley Wireless Commun. Mobile Comput. (WCMC) J. 8 (2008) 171–193.

[16] Girao, J.; Westhoff, D.; Mykletun, E.; Araki, T. TinyPEDS: Tiny persistent encrypted data storage in asynchronous wireless ad-hoc networks. Ad Hoc Netw. 2007, 5, 1073–1089.

Approaching Machine Learning Using H2O and Inspecting it on Apparel Industry Using D3 Visualization

Chaitra D.B, Mrs.Bindiya M.K
PG SCHALOR, Associate Professor
Dept. of Computer Science and Engineering
SJB Institute of Technology Bangalore-60, India
chaitradbhogesh@gmail.com, bindiyamk2004@yahoo.com

ABSTRACT- Implementing machine learning algorithms was never a easy task, Although there are several approach to do so, but the procedure followed is always exasperating. In order to simplify the task of analyses with same accuracy we give a new approach of H2O.H2O is fast ,scalable, distributed, machine learning and deep learning application .It is a smarter Which implements many generalized linear models like linear regression, logistic regression ,Naive Bayes, K-means clustering, Naive Bayes algorithms .Here in this paper we are trying to approach H2O machine learning with more complex algorithms with an effective solution which can be which can be implemented on various real time problems, one such among them is apparel industry which still follows a traditional inspection and analysis system so using machine learning and D3 visualization we are trying to give better analyses and recommendation to the system.

Keywords -Machine Learning, H2O, Analytics, GLM algorithm, D3 visualization.

I. INTRODUCTION

Over the past few decades Machine Learning has become one of the important part of technology. With the ever increasing amounts of data and which gives rise to a good reason to believe that smart analysis will become even more necessary ingredient for the technological process.

Machine Learning deals with automating the automation that is making the machine intelligent [1].Using the past example data or using the past experience to solve a given problem many successful applications have been developed one such application is H2O [2].

A. H2O Machine Learning

H2O is one of the machine learning application which makes it easy for anyone to apply predictive analysis and math to solve today's any of the most challenging business problems. Many features of machine learning have been intelligently combined this platform which is nor currently present in any other machine learning platforms.

In H2O it is easy to combine the power of different highly advanced algorithms and the truly scalable in memory processing capacity for big data to increase the benefit for the business [3]. With H2O it is possible to make better predictions, ready to use algorithms can be harnessed and also the super power which is needed to analyze bigger data sets, more variables and models can be easily obtained. With H2O we can work with any of existing languages and also tools and can even expand the platform seamlessly to our Hadoop environments.

H2O uses familiar programming environments like Python,R, JSON,Scala and JAVA through H2O's powerful API's.Using H2O we can explore the model of any big data from within R Studio,Microsoft Excel and Tableau and many more.Through H2O it is easy to connect data from

NoSQL, S3, HDFS and SQL and other data sources and H2O can be installed on any platform and deployed also.

H2O supports Nano Fast scoring engine and also we can train any number of data models and can iterate over those models to accurate models in real time using H2O's in memory distributed parallel processing develop.

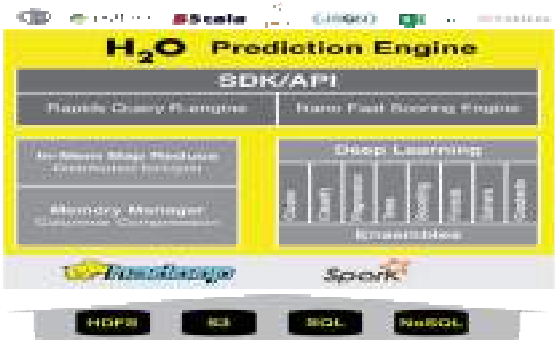


Figure 1: H2O Machine Learning Architecture

Once the analysis of the data has been done using the H2O machine learning we do the visualization of the data using the D3 [4].

B. D3 visualization

D3 gives us the liberty to bind any arbitrary data to a DOM that is Document Object Model and then we apply the data driven transformations to those documents [5].D3 a data driven document is just not a monolithic framework that only provide every conceivable feature, but also solves the crux of the given problem and does the efficient manipulation of the documents based on the data contained in that particular document. This quality of D3 avoids the

proprietary representation of data by enabling the extraordinary flexibility and also exposes the web standards like SVG,HTML and CSS .D3 is a functional style representation which can be modified according to the needs.D3 with the little overhead is extremely fast and it supports any range of large datasets and most important thing of D3 is that it gives the dynamic interaction to the graphs and animations because of which richness is added to the D3 visualization .

II.RELATED WORK

Work on "Social Contextual Recommendation" that there is an epidemic growth in the data by social media which is cannot be handled by the traditional database systems. The existing tools consider the social structure of network but not the social context in full so recommend the things hence he proposes a model which does the recommendation based on the sociology study of the person on his different social media activities and his individual likes and dislikes his preferences to do the recommendations. On the basis of his interpersonal influences also recommendation is done. The empirical results are based on the datasets collected from Facebook, Twitter.

Paper "Layer-Cantered Approach for Multigraphs Visualization" he demonstrates that recent advances in the social media, network sciences involves the creation of models and analysis of generated data. . In this paper he demonstrates the new model for multiple edge graphs and to manage the different types of relationship between existing multigraphs. In this approach the pair of edges or nodes are specifically studied, and how the different nodes are interacted together based on the edges they share. In this paper in two level of global and local multigraph features are considered. The global approach deals with the gaining the knowledge which are related to the different characteristics of layer they combine with, where in the local gives the analysis of separate layers and explain each characteristics separately. This proposal is tested with real world data and it has helped to bring about useful patterns.

III. MOTIVATING EXAMPLE

Big data analytics is picking up a large data sets of high volume and high variety of data because of which it is called big data and processing it for business intelligence. There are many customer preferences, hidden patterns, unknown correlations and also many other business information which are useful for analysis.

In many social media websites they use re commander systems to recommend the new movies, songs, books and many other things based on the correlation of the members in the social media .If two persons like something in common along with their other personal unique likes system tries to recommend those unique likes also to each individual who share few things in common thinking that they may like these recommendations also. This is how the market promotion is done for items these days .Association rules also plays important roles in this recommendation systems like if someone buys phone he may be recommended with memory cards likewise.

Using the quite a same is expanded to apparel industry ,there are various band apparels now a days all the apparels manufacturer does not need to own his own industry he may give his orders to other local industries who take the orders from any customers and develop their needs and supply to them.

The interesting part of this process is inspecting the developed products from the suppliers by the inspectors .There may be any kind of mistakes from the suppliers end like the material of the cloth may be different or accessories used, quality of the product, size of the product likewise there can be any number of defects or even it may be perfect .The quality manager or the inspector may find and need to generate report. Thereby huge number of reports are being generated across the globe.

Thus we have taken all these reports and analyzed for the quality of each supplier to each customer for all possible years and found who can be the most reliable supplier ,what are his defects rates, is that increased or decreased compared to older years and predicting about its future years rate. By taking his overall performance we can recommend to the customer whether it's worth giving order to that particular supplier or not. All this analyzed data is represented in eye catchy D3 visualization techniques so that it's easy of common people to understandable and consider the recommendation.

IV. THE APPROACH

We approached the problem of apparel industry through the concepts of machine learning and used other technologies like NO-SQL, MongoDB, JSON, SQLite and also visualization technique called D3.We have approached machine learning here using H2O platform. First considering the reports generated by the inspectors ,collected all the reports in which the format they are generated that is either in the form of CSV ,Microsoft excel or in any other format. The data inside these documents is the input data and this data is used for the training the systems.

The reports irrespective of the form they are generated they are they are converted into standard JSON which is first manually created and checked for the universality for all other reports and once confirmed it is automated .Any new report comes it automatically converted into the JSON structure defined .Then the dump of SQL tables are generated using which JSON files are created ,for every report there is single file generated for the purpose of extracting the required data and then moving the files to MongoDB which is a NO-SQL database for the further process of analysis.

Once the data is available on the MongoDB it passed on to the H2O machine learning platform using R. Then the data is analyzed using the GLM (Generalized Linear

Model) which is the algorithm for both classification and regression using which when a new supplier for the customer comes in it decides whether that supplier is the high ,low or moderate customers for that customer that is classification is done .Then after analyzing all the data the visualization is done using the D3 graphs these are the highly interactive graphs wherein a person can easily understand the analysis done without missing the single part.

A.The GLM algorithm

GLM is one of the flexible model in statistics, it is the generalization of the ordinary linear regression model. GLM is for both classification and regression [6]. GLM estimates the regression for the exponential distribution values. GLM along with the Gaussian distribution it also supports Poisson, Gamma, Binomial and also Tweedie distributions

Generalized GLM is given by relating the linear model via a link function and by taking the range of the variance for each measurement to be a function of its predicted value. GLM allows response variables that have error distribution other than normal distribution [7] [8] [9].

For the estimation of the model ,an iteratively reweighed least square method is considered for maximum likelihood .GLM were form as a way of unifying other statistical models like logistic regression, linear and poisson regression. Least square fits and Bayesian approaches to variance stabilized responses also have been developed.

B.The GLM suite includes

- Gaussian regression
- Gamma regression
- Poisson regression
- Tweedie regression and
- Binomial regression

C. Usage

h2o.glm(x, y, data, key = "", offset = NULL, family, link, tweedie.p = ifelse(family="tweedie", 1.5),strong_rules = TRUE, alpha = 0.5, prior = NULL, standardize = TRUE,beta_constraints = NULL, nfolds = 0 , use_all_factor_levels = FALSE, lambda_search = FALSE, , disable_line_search = FALSE,nlambda = -1, max_predictors = -1, return_all_lambda = FALSE, intercept = TRUE,lambda.min.ratio = -1, non_negative = FALSE,variable_importances = FALSE, iter.max = 100, higher_accuracy = FALSE)

D.Variables of GLM Model

Response: It is the model dependent variable and is denoted as Y .While choosing the appropriate model for distribution for estimating specific features of a dependent variable should be taken into consideration

Gaussian: Y must be real valued and contiguous variable

Gamma: Y variable must be valued strictly greater than 0 and also it must be discrete

Binomial: Y variable are valued only at 0 or 1 and also discrete

Tweedie: Y variable are combination of Poisson-Gamma compound distribution

Poisson: It is used as a model count data. These Y variables are strictly greater than 0.

Ignored Columns: From the current data set a list of columns field will be auto-populated.H2O gives the liberty of selecting the set of columns which can be omitted from the processing model.H2O by itself omits the column with the constant values and also dependent variable in Y. Since variances are constant columns are omitted.

Standardize: Variables are transformed into standardized variables, each with unit variance and mean value of 0.Co-efficients and Variables are expressed with respect to their standard units and relative 0 position.

Maximum iteration: When the data set is given to the system, these datasets are iteratively processed to increase the accuracy. As the number of iteration increases accuracy also increases. If maximum iteration is set to 100 then algorithm repeats the gradient descent 100

times .These iterations are performed for training examples. N folds, this is the number which specifies cross-validation models to generate parallel for training a model on the whole data set. If value of Nfolds is set to 10, additional models are generated with 1/10 of data for each data.

Family and Link: In H2O GLM function each family is connected with a default link function, each function defines the specialized transformation on the set of variables that is X variables chosen to predict Y.

Gaussian: It is the identity here the Y are contiguous or discrete predicted values and quantitative that can be interpreted as approximately continuous.

Gamma: It is the inverse, and even distributed as Poisson, here the variance is greater than mean.

Binomial: It is the logit value. These binomial dependent variables are taken as categorical Y with two possible outcomes 0 and 1.

E.Validate GLM:

Cross Validation: GLM analyses in H2O are presented with cross-validation models, here the coefficients

presented in the model are not dependent on the cross-validation models, and the coefficients are generated with least-squares on the whole dataset. Cross validation values are generated by taking 90% of sub sample data and remaining 10% with training and test data [10].

Cost of Computation: H2O gives a distributed parallel computing hence large amount of data can be processed, here the large sets are divided into smaller sets and then processed [11].

In GLM, data are divided into rows and not columns, this is because the Y value are dependent on information of predicator variable vectors.

If O is taken as complexity function and P as number of columns or predictors and N as the number of rows or observation then

Equation 1: $$Runtime \propto p^3 + \frac{(N * p^2)}{CPU_s}$$

By above formula it is evident that distribution reduces the time it takes to process the algorithm as it decreases N.

F.GLM Algorithm

Let $y_1, \cdots \cdots y_n$ be n observations of the independent, random response variable y_i [13] [14].

Assume that the observations are distributed according to a function from the exponential family and have a probability density function of the form:

Equation2: $$f(y_i) = \exp\left[\frac{y_i^{\theta_i} - b(\theta_i)}{a_i(\phi)} + c(y_i;\phi)\right]$$

Where θ and ϕ are location and scale parameters. And $a_i(\phi)$, $b(\theta_i)$, $c(y_i;\phi)$ are known functions. a_i Is of the form: $a_i = \frac{\phi}{p_i}; p_i$ is a known prior weight. When Y has a pdf from the exponential family:

$$\vee(\theta_i)a_i(\phi)$$

$$E(Y_i) = \mu_i = b \; var(Y_i) = \sigma_i^2 = b$$

Let $g(\mu_i) = \eta_i$ be a monotonic, differentiable transformation of the expected value of y_i. The function η_i is the link function and follows a linear model. $g(\mu_i) = \eta_i = x_i^{\prime}\beta$

When inverted: $\mu = g^{-1}(x_i^{\prime}\beta)$

G .Other technologies used

Mongo DB: Mongo DB is one of the widely used NoSQL database .It is cross-platform and document oriented document database. MongoDB is widely used since it entangles the traditional table based data that is relational database in the favor of JSON like documents.

MongoDB has features like AD hoc queries, Load balancing, Replication, File storage, importantly aggregation and many other.

It also gives the high performance, availability and automatic sharding. The records are stored in the form of key and value pairs. The value of the filed maybe any other documents or arrays or even the arrays of documents also. Since all the data are in form of objects it is easy to access and analyze .It supports horizontal scalability and also multiple storage engines.

JSON: It is the abbreviated form of JavaScript Object Notation, as the name itself specifies the data are stored in the form of key and value pairs and values can be again a document or arrays. It is very easy for humans to read and write this kind of format .It is in the text format and fully independent of the language.

JSON is built in two ways, one is a collection of key and value pairs like record, object, associative array and other is ordered list of values like vector, array, list etc. JSON is a unordered set of data. A data is embraced in open and close braces.

V. ALGORITHMS

A. Algorithm for Creating JSON

Step 1: Gathering the required dump from client, dump contains the data of inspected reports which can be in any format.

Step2: Converting the dump into SQL tables using SQLite

Step3: Processing the tables to generate JSON, for each data separate JSON

Step4: Cleaning the JSON and pre-processing for convenient or required data.

Step5: Loading the obtained resultant JSON to MongoDB

B. Algorithm for processing the data inside MongoDB

Step1: Create an API for data extraction from JSON in MongoDB

Step 2: Process the extracted data into machine readable JSON format output

Step3: Pass the JSON output to the GLM algorithm for analysis

Step 4: compare the results of all the datasets previous and current if produced result is acceptable then accept else repeat step 2 and 3

Step 5: once output is acceptable pass it to D3 charts for visualization

C. *Algorithm for Visualization*

Step 1: Pass the output result of GLM algorithm to D3 charts for generating graphs

Step 2: Collect feedback from D3 charts to make it interactive, if interaction works correctly it is accepted or else change the blocks used for interaction

VI. EXPECTED RESULTS

In this section we demonstrate the expected output of our approach alongside comparing the existing model results. Specifically, we demonstrate on working of algorithm: 1) creating the JSON 2) loading data to the database 3) analyzing the data using H2O and 4) visualization using D3 chats.

A. Creating JSON:

In the creation of the JSON here we create one single standard JSON .This JSON structure is common for all reports data any data coming in will be converted to this standard JSON format which is in the form of key and value pairs which is easy to understand and access for further analyses .Whereas the existing model contains the report in the way they are generated like Microsoft Excel ,CSV of any other format which is difficult to access and analyze and get the relevant result .Different reports are in different manner and it is tedious to match the compatibility of the reports and because of which the accuracy of the result goes down. Hence the conversion of all the reports to single standard format helps to obtain the fast processing and also accurate result.

B. Moving JSON to MongoDB:

After creating the standard JSON structure it is moved to the No-SQL database MongoDB. MongoDB is a schema less database because of which we can take a liberty of changing the report formats .The data coming in may contain the missing data ,redundant data or even the data which is not useful for the process ,taking all this parameters MongoDB is the best database as it by default gives all these features and aggregation is done which is required for the analyses .MongoDB gives the best aggregation queries to be used and in our result it gives the various aggregation result like according to the aggregating the suppliers and customers which supplier quantity of the items supplied all these aggregations are easily done again which is very much required for analyses process. Whereas which is not possible in existing system which uses the relational database Which is schema dependent database and using which dynamic schema cannot be generated.

C.Analysis:

After passing all the data to the MongoDB, considering only relevant data analyses is carried out. Data contains the information about the client ,supplier, location , quantity of items built and supplied to the customer and what are the major defects ,minor ones and which needs to be prioritized first and all the analytics is done here all these analyses is done using the GLM algorithm and this algorithm gives the accurate result. By using the results of the analyses one can find the moderate, average and low suppliers and based on this analyses companies can decide to which supplier it can consider next. Where there is no analyses done in the present system.

D. D3 Visualization:

Once the analyses result is got it can be visualized using D3 chats which are very interactive and easy to understand the result. These interactive graphs gives the clear picture of overall analyses done about client, supplier, quality ,quantity, their consistency range over the years .Here we are generating the stacked bar chart ,Sequential Sunburst, Pie charts and line charts.

In the stacked bar graph we compare the supplier and the quantity of items he has delivered to his customers over a period of time. This gives the consistency analyses of the supplier over range of years. This graph is made interactive by taking either staked or grouped result.

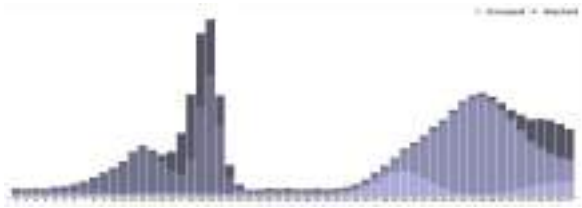


Figure 2: Stacked bar graph

Using the pie charts we represented who is the low ,medium and high suppliers over a period of time .Here based on the percentage of the mid, high and low pie chart is distributed .It is made highly interactive by clicking on one area of pie charts it gives pop of another bar chart with relative values.

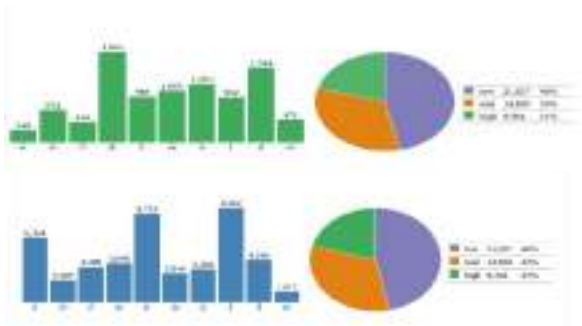
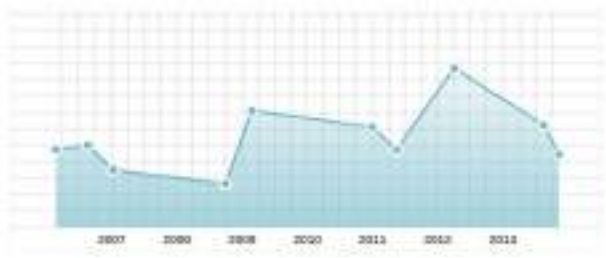


Figure3 a) Pie chart for low range b) Pie chart for mid-range



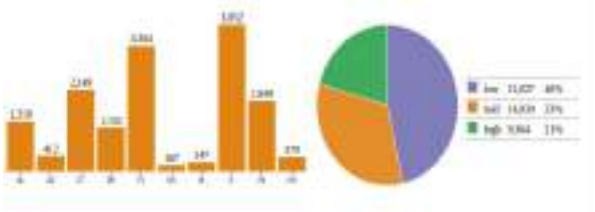


Figure 4: Pie chart for high range

Area graph is used to give the performance of the particular supplier over a period of time or years.

Figure5: Area graph over range of years

Overall analyses can be even represented in a D3 chart called sequential sunburst. This graph gives the information about the client, supplier, location of the industry, name of the industry and all other values. By clicking on one part of the chart it redraws and takes the other values according to

the requirement.

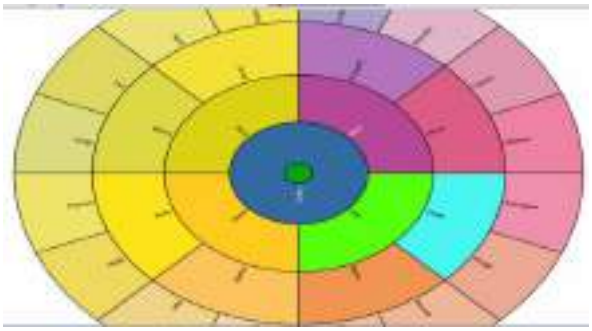
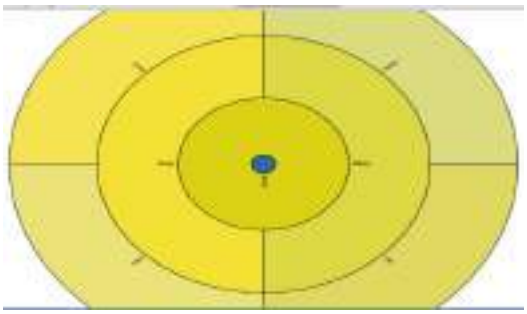


Figure6: Sequential sunburst for overall model



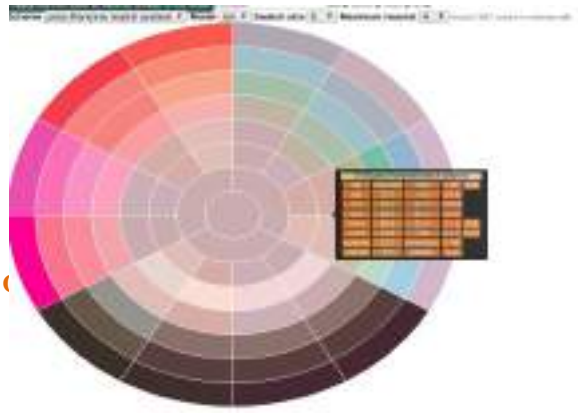


Figure7: Sequential sunburst when one level reduced

Figure8: Sequential sunburst when two level reduced

Figure9: Sunburst with popup table information

CONCLUSION

In this paper we have tried to analyze the data of apparel industry and give them a better approach to analytics using H2O machine learning application. We have tried to convert all the reports to one single standard structure of JSON and successfully stored to MongoDB and from where the analysis is done for the better supplier and customer and quality of the product which is recommended for the people and also future predictions and customer performance is found. At the last all the analysis is represented in the form of charts for the better understanding. Bylooking into the graph one can predict how is the moderate, low or high supplier and also customer, quality of the product, quantity of the product supplied and what can be his future quantity and quality and where is the location of the industry, who are all the customer for which customer and other possible analytics of all the apparel industry present across the globe. The results obtained were better promising then the existing model.

REFERENCES

[1] http://www.sas.com/en_us/insights/analytics/machine-learning.html
[2] https://www.r-project.org/about.html
[3] http://h2o.ai/product/algorithms/
[4] http://searchbusinessanalytics.techtarget.com/definition/data-visualization
[5] https://d3js.org/
[6] http://h2orelease.s3.amazonaws.com/h2o/master/1713/docs-website/datascience/glm.html
[7] Breslow, N E. "Generalized Linear Models: Checking Assumptions and Strengthening Conclusions," Statistica Applicata 8 (1996): 23-41.

[8] Frome, E L, "The Analysis of Rates Using Poisson Regression Models," Biometrics (1983): 665-674. http://www.csm.ornl.gov/~frome/BE/FP/FromeBiometrics83.pdf

[9] Goldberger, Arthur S, "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," Journal of the American Statistical Association 57.298 (1962): 369-375.

[10] Guisan, Antoine, Thomas C Edwards Jr, and Trevor Hastie, "Generalized Linear and Generalized Additive Models in Studies of Species Distributions: Setting the Scene," Ecological modeling 157.2 (2002): 89-100. http://www.stanford.edu/~hastie/Papers/GuisanEtAl_EcolModel-2003.pdf

[11] Nelder, John A, and Robert WM Wedderburn, "Generalized Linear Models," Journal of the Royal Statistical Society. Series a (General) (1972): 370-384. http://biecek.pl/MIMUW/uploads/Nelder_GLM.pdf

[12] Niu, Feng, et al. "Hogwild: A lock-free approach to parallelizing stochastic gradient descent," Advances in Neural Information Processing Systems 24 (2011):693-701.http://www.eecs.berkeley.edu/~brecht/papers/hogwildTR.pdf

International Journal of Advanced Networking and Applications (IJANA)

ISSN : 0975-0290 (Print) **ISSN : 0975-0282 (Online)**

Greetings…

The **IJANA** *is a* bi-monthly Journal published by Eswar Publications focusing papers related to networking and its applications since 2009.

Impact Factor of the Journal is : 3.462

Index Copernicus Value ICV 5.36 points.

The journal is welcomed and supported by the academicians, researchers and students from various parts of the Globe. The journal is already indexed by the following abstracting / indexing services

To make your paper published in one of the standard journal, we welcome the unpublished paper from the authors.

Send your unpublished paper for the consideration in the forthcoming issues to **editor.ijana@gmail.com**.

Have a nice day.

With regards
Editorial Team, IJANA
Web : www.ijana.in
Email: editor.ijana@gmail.com