

Amity University
UTTAR PRADESH



INTERNATIONAL CONFERENCE ON TELECOMMUNICATION AND NETWORKS (TEL-NET 2013)



27-28 FEBRUARY, 2013

Organized By

**Amity Institute of Telecom Technology
and Management (AITTM)**

&

IETE Sub-Centre, Amity University Uttar Pradesh, India

In association with

**The Institution of Electronics
and Telecommunication Engineers (IETE)**

**INTERNATIONAL CONFERENCE
ON
TELECOMMUNICATION AND NETWORKS
(TEL-NET 2013)**

**International Conference
on
Telecommunication and Networks**

TEL-NET 2013



27th - 28th February, 2013
Amity University, Uttar Pradesh, Noida, India

Organized by



Amity Institute of Telecom Technology and Management (AITTM)
&
IETE Sub-Centre, Amity University, Uttar Pradesh, Noida, INDIA

in Association with
The Institution of Electronics and Telecommunication Engineers (IETE)

First Impression: 2013

© Amity Institute of Telecom Technology and Management,
Amity University, Uttar Pradesh, Noida, India

**International Conference on Telecommunication and Networks
TEL-NET 2013**

ISBN: 978-93-81583-93-7

No part of this publication may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owners.

DISCLAIMER

The authors are solely responsible for the contents of the papers compiled in this volume. The publishers or editors do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

Published by

EXCELLENT PUBLISHERS

Kishangarh, Vasant Kunj, New Delhi-110 070

Tel.: 9910948516, 9958167102

E-mail: exlpubservices@gmail.com

Typeset by

Excellent Publishing Services, New Delhi-110 070

Preface

1. Telecommunications has been recognised world over as an important driver for socio economic development of a nation. In India, the unprecedented growth in the tele density in the last decade and half has contributed 3 % growth in the GDP. The World Bank has estimated that every 10 % increase in the Broad Band penetration in the developing countries increases the GDP by 1.38%, which is a substantial amount. India has realized the potential of the proliferation of Information Communication Technologies and has vision to transform the country into a knowledge based society using Telecommunication as the platform. Telephone today is not a mere communication device but an instrument of empowerment.
2. Digitization of telecommunication networks has enabled transmission of voice, video and data on the same network. The user aspirations and requirements are putting more and more demands on the telecommunication networks in terms of higher speeds, more bandwidths and better Quality of Service. The speeds which were considered adequate few years ago do not even qualify for providing most basic services now. The networks today are required to be flatter, faster and smarter which discover the optimal paths from point to point ensuring zero packet loss.
3. The telecom technologies are fast evolving both at the access networks and the core networks. The 4G mobile wireless technologies such as WiMax, LTE, and advanced LTE promises similar data rates as available on fixed networks. The FTTH technology provides converged services i.e voice, data, video, IP TV, video on demand, internet services etc. on a single fiber to users at their homes. The core networks are being engineered on the self healing optical networks with unlimited bandwidth and fast switching using Multi Protocol Label Switching are ensuring that the delivery of data is error free at the required speed. These emerging telecommunication technologies are throwing up new challenges in terms of availability of spectrum, radiated power, network and information security, power consumption, data speeds, reliability of service etc. The biggest challenge however remains in development of the human resource with integrated skills for the converged ICT sector. India has an advantage due to young populations and can drive maximum dividend from their creative abilities. This Conference therefore aims to explore growing advancements in the fields of Telecommunication and Networking Technologies, and provides a common platform to leading scientists, academicians, researchers, government officials, practicing engineers, industry professionals and students to share their research experiences and views. The conference will focus on the following thrust areas:
 - Wireless Communication Technologies
 - Emerging Telecom Technologies
 - Optoelectronics and Optical Communication
 - Ad Hoc and Sensor Networks
 - Antennas and Wave Processing
 - Image and Video Processing
 - Network Security
 - Electronics for Telecommunication

Acknowledgement

We wish to express our appreciation and sincere gratitude to Dr. Ashok K. Chauhan, our beloved Founder President, Ritnand Balved Educational Foundation (RBEF) (Foundation of Amity Institutions & Sponsoring body of Amity Universities), Mr. Atul Chauhan, Chancellor, Amity University, Uttar Pradesh (AUUP), Prof. (Dr.) Balvinder Shukla, Vice Chancellor (Actg.) AUUP, Lt. Gen. P.D. Bhargava, Group Deputy Vice Chancellor and Advisor AITTM, AUUP for their constant encouragement and guidance during the planning and conduct of TEL-NET 2013 Conference from the proceedings of which this book has evolved.

This Conference has been organized in association with The Institution of Electronics and Telecommunication Engineers (IETE). We would like to acknowledge support extended by Dr. Surendra Pal, President – IETE, Dr. M.H. Kori, Governing Council Member, IETE, Shri S.K. Aggarwal, Secretary General IETE, Shri S.K. Arora, Public Relations Officer and complete staff of IETE.

We will like to thank our esteem members of International Advisory Committee and National Advisory Committee. We gratefully acknowledge the Visionary speakers, foreign delegates, authors of research papers and participants who have encouraged us by participating in large number.

We will like to extend our special thanks to Prof. R.K. Shevgaonkar, Director IIT, Delhi, Mr. Vimal Wakhlu, CMD TCIL, Mr. M.I. Bhat, Project Director-ICT Sector Development Project of MCIT Afghanistan, Prof. (Dr.) Shibani Koul, Professor & Head, CARE, IIT Delhi, Dr. Rajeev Shorey, Founder President and Advisor, NIIT University, Nimrana, Prof. S.P. Ojha, Former Vice Chancellor, CCS University, Mr. Rajat Mukarji, Chief Corporate Affairs, Idea Cellular, Prof. (Dr.) D.R. Bhaskar, Jamia Millia Islamia University, New Delhi, Prof. R.K. Sinha, DTU, New Delhi, Mr. S.N. Gupta, Chief-Corporate Affairs, Sterlite Technologies Ltd, New Delhi, Former Member of TRAI, Mr. Mahesh Khera, Founder Director, KTMT Consulting, India, Col. KPM Das, Senior Vice President, CISCO, Prof. (Dr.) Sunil Kumar Khatri Director AIIT, Amity university, Ms. Chhaya Chordia, Jt. Director Hostel, Amity University, Administrative department of Amity University, Members of Organizing Committees, for whole hearted support for the conduct of the Conference.

We also acknowledge moral support extended by the Head of Institutions / Departments / Independent Centers of Amity University Uttar Pradesh for the Conference.

Publication of this book within time would not have been possible without the active support of our publishers, Excellent Publishing House, New Delhi.

Committees

PATRONS

Dr. Ashok K Chauhan

Founder President, Ritnand Balved Educational Foundation
(RBEF) (Foundation of Amity Institutions
& Sponsoring body of Amity Universities)

Sh. Atul Chauhan

Chancellor
Amity University Uttar Pradesh (AUUP), India

VICE-PATRON

Prof. (Dr.) Balvinder Shukla

Acting Vice-Chancellor, AUUP, India

CONFERENCE CHAIR

Lt Gen. P D Bhargava

Deputy Group Vice-Chancellor, AUUP, India
Advisor, AITTM

CONFERENCE CO-CHAIR & CONVENER

Prof. R K Kapur

Dy. Director & Head, AITTM, AUUP, India

TECHNICAL CHAIR

Dr. M.H. Kori

Governing Council Member, IETE
Technology Consultant,
Technology Advisor, Validus Technologies USA,
Retd. Technical Director, Alcatel-Lucent Technologies

Organizing Committee

Secretary

Mr. Arvind Kumar, AITTM, AUUP, India

Treasurer

Mr. Anil Kumar Sajnani, AITTM, AUUP, India

Editorial Committee

Mr. Anil Kumar Shukla, AITTM, AUUP, India

Ms. A. Devi Priya, AITTM, AUUP, India

Mr. Atul Kumar Srivastva, AITTM, AUUP, India

Ms. Manisha Manoj, AITTM, AUUP, India

Conference Coordinating Committee

Mr. J. S. Jadon, AITTM, AUUP, India

Dr. Dheeraj Pawar, AITTM, AUUP, India

Dr. Arun Kumar, AITTM, AUUP, India

Ms. Manisha Gururani, AITTM, AUUP, India

Ms. Shubra Dixit, AITTM, AUUP, India

Ms. Neha Arora, AITTM, AUUP, India

Hospitality Committee

Prof. P. S. Bajaj, AITTM, AUUP, India

Mr. V. K. Sharma, AITTM, AUUP, India

Mr. Mukul Varshney, AITTM, AUUP, Noida

Mr. R.C. Singh, AITTM, AUUP, India

Secretariat Support

Ms. M. Vijaya, AITTM, AUUP, India

Ms. Bakul Sharma, AITTM, AUUP, India

Speakers

Prof (Dr.) R. K Shevgaonkar

Director, IIT, Delhi

Dr. Surendra Pal

President,
The Institution of Electronics
and Telecommunication Engineers
(IETE), India

Mr. Vimal Wakhlu

Chairman & Managing Director,
Telecommunications Consultants India Limited
(A Government of India Enterprise)

Dr. M. H. Kori

Governing Council Member, IETE
Technology Consultant,
Technology Advisor, Validus Technologies USA,
Retd. Technical Director, Alcatel-Lucent
Technologies

Dr. Rajeev Shorey

Founding President and Advisor
NIIT University, Nimrana

Prof. (Dr.) Shibani Koul

Professor & Head
Centre for Applied Research in Electronics
CARE, IIT Delhi

Mr. Rajat Mukarji

Chief Corporate affair,
Idea Cellular Pvt. Ltd.

Prof. S. P. Ojha

Former Vice-Chancellor, C.C.S. University,
Meerut

Mr. Mahesh Khera

Founder KTMT, India
India Chief of Strategy and Business
Development, Location Labs, CA,US Premium
Partner of Global Consulting Platform Providers,
Member Technology and Regulatory Advisory
Committee (TRAC), IETE, India

Mr. Satya N. Gupta

Chief - Corporate Affairs,
Sterlite Technologies Limited
New Delhi, India Govt. Of India,
Former Member TRAI

Prof. D. R. Bhaskar

Professor, Jamia Millia Islamia University,
Member NAAC

Prof. Ravindra Kumar Sinha

Delhi Technical University,
Bawana Road, Delhi-110042

Dr. Mihir Mohanty

Associate Prof. Dept of ECE ITER, Orissa

Mr. V.K Arya

Deputy General, Telecommunications
Engineering Center
Dept of Telecommunications, Ministry of
Communications & IT, Govt of India,

National Advisory Committee Members

Dr. Surendra Pal

President
The Institution of Electronics and
Telecommunication Engineers (IETE)

Prof. Zahid Husain Khan

Director: FTK-Centre for Information
Technology, Jamia Millia Islamia University,
New Delhi.
Vice-President: Delhi Education Society.
Member, National Mission on Education
through Information Communication Member
Technology and Regulatory Advisory
Technology (NMEICT, Govt. of India).

Mr. Satya N. Gupta

Chief - Corporate Affairs
Sterlite Technologies Limited
New Delhi, India

Dr. S. K. Aggarwal

Former -Director e-infrastructure projects
Government of India ,DEITY, India
(Council member of IETE)

Prof. Mihir Narayan Mohanty

Department of Electronics and
Communication
Engineering, Institute of Technical Education
and Research SOA University, Bhubaneswar

Prof. A K Sharma

Dean, HOD(E&C), YMCA, Faridabad

Dr. Puneet Sabbarwal

Senior DFT Design Engineer Texas
Instruments

Mr. Arif Shouqi

Chief Defense Architect
Asia Pacific Region
Cisco

Col. O P Arora

Advisor IIC, AITTM, AUUP,NOIDA

Dr. M. H. Kori

Technology Consultant,
Technology Advisor, Validus Technologies USA, Retd
Technical Director, Alcatel-Lucent Technologies

Mr. Mahesh Khara

Founder KTMT, India
India Chief of Strategy and Business Development,
Location Labs, CA, US
Premium Partner of Global Consulting Platform
Providers

Mr. Setumadhavan

Director, Strategy and Marketing HuaWei

Prof. Sunil Kumar Khatri

Head AIIT, AUUP, Noida

Mr. V K Arya

Senior General Manager, BSNL
Dept of Telecommunications, Ministry of Communications
& IT, Govt of India

Prof. Vivekanand Mishra

SVNIT, Elec. & Comm. Department, Gujrat
Senior Member IEEE

Mr. Ramesh A Aditya

Director- Training and Placements
SKP Group of Institutions
Tiruvannamalai

Prof. Ravindra Kumar Sinha

Head of the Department, Delhi Technological University
(Formerly Delhi College of Engineering, University of
Delhi), Chief Coordinator: TIFAC-Centre of Relevance and
Excellence in Fibre Optics and Optical Communication,
Mission REACH Program, Technology Vision-2020, Govt.
of India

International Advisory Committee Members

Prof. R K Shevgaonkar

Director
Indian Institute of Technology(IIT),
Delhi, INDIA

Prof. Ashok Jhunjunwala

Indian Institute of Technology (IIT), Madras.
Leader: Telecommunications and Computer Networks
group (TeNeT), IIT Madras. INDIA
Member of Prime Minister's Setup Scientific Advisory
Committee

Mohammad Ismail Bhat

Head of PMO and the Project Director of
World Bank funded ICT Sector Development
Project of Afghanistan, Ministry of
Communications and IT
KABUL-AFGHANISTAN

Mr. Joshua McCloud

Customer Solutions Architect
Cisco Asia Pacific, Japan, China, Public Sector

Mr. Johan Lindgren

CEO, Eniro Group; a Local Search
Company Operating in Sweden, Norway,
Denmark, Finland and Sweden

Mr. Adrian Topp

Head of MS Network Operations
Ericsson India Global Services Pvt, Limited

Prof. D.E. Ventzas (SMISA)

TEI of Larissa
Department of Informatics and
Telecommunications
GREECE

Prof. F. Feng

Xi'an University of Posts and
Telecommunications Xi'an, China

Dr. Andreu Vea

Internet Society (ISOC-ES)
President of the Board
World Summit Awards (UNESCO)
Eminent Expert for Spain

Prof. Huzur Saran

Indian Institute of Technology (IIT), Delhi, *INDIA*
Department of Computer Science &
Engineering

Prof. Stefka Stoyanova Fidanova

Institute of Parallel Processing Bulgarian
Academy of Science, Sofia
Bulgaria

Dr. Ujjwal Kumar

KNMI - Royal Netherlands Meteorological Institute
De Bilt, Netherlands

Prof. S. P. Ojha

Former Vice-Chancellor, C.C.S. University,
Meerut, INDIA

Ms. Samta Suman

Dallas, Texas, United State

Prof. K. K. Aggarwal

Chancellor, Lingaya's University, Faridabad,
Haryana, INDIA
Ex-Vice Chancellor, GGSIP University

Founder President's Message



I am delighted to learn that Amity Institute of Telecom Technology and Management (AITTM) and IETE Sub-Centre Amity University UP, NOIDA are organizing International Conference on Telecommunication and Networks (TEL-NET 2013) in association with Institution of Electronics and Telecommunication Engineers (IETE).

Telecommunication Networks are the basic infrastructure for the accelerated, equitable and inclusive growth of the nation. India is the second largest market for the Telecom Industry and in next five years, it is expected that an investment of approximately 1.5 trillion rupees will take place in this sector. It will require huge trained human resource. I am proud that AITTM, Amity University UP, is providing since long years Industry ready engineers and technocrats for Telecom Industry.

I am sure this conference will provide a common platform to experts in the core field, researchers and industry captains to exchange their views on the changing paradigm in the field of Telecommunication and Networks. It will provide an opportunity to all participants to widen their domain knowledge and ignite their minds to undertake new research in this field.

I appreciate the vision of President and other office bearers of Institution of Electronics and Telecommunication Engineers (IETE) for establishing IETE Sub-Centre at Amity University which is definitely going to become among the best sub-Centre of IETE in the country in times to come.

I extend a very cordial welcome to all worthy speakers and capable participants from India and abroad, and wish them fruitful deliberations and a pleasant stay.

I greatly appreciate Lt Gen P D Bhargava, the Conference Chair and Group Dy Vice Chancellor & Advisor - AITTM, Professor Raj Kamal Kapur, Convenor Tel - Net 2013, Dy. Director & Head – AITTM and their dedicated team for their praiseworthy efforts in ensuring the success of this Conference. Also, the guidance, advice and support provided by Prof Surinder Pal, President IETE needs special mention.

I appreciate all the sponsors and partners and other organisations whose valuable contribution will surely lead to achieving long-term sustainable results in this area.

I hope that the Conference will provide new opportunities for collaboration and would define the road-map for undertaking future valuable activities in this field.

I wish the Conference a grand success.

Dr Ashok K Chauhan
Founder President, Ritnand Balved Education Foundation (RBEF)
(The foundation of Amity Institutions and
the Sponsoring Body of Amity Universities)

Message from Chancellor Amity University Uttar Pradesh



I am happy to know that Amity Institute of Telecom Technology and Management (AITTM) and IETE Sub-Centre Amity University UP, NOIDA, are organizing International Conference on Telecommunication and Networks (TEL-NET 2013) in association with Institution of Electronics and Telecommunication Engineers (IETE).

Telecommunication and Information technologies have changed the socio economic environment in last two decades. The older technologies are being replaced by newer generation of technologies with amazing features which has changed the way we interact with each other. These technologies have become more user friendly but at the same time the complexities have increased many times for the operators. Our students and researchers have not only to keep abreast with emerging technologies but also device new technologies which are environment friendly and use scarce resources of spectrum, power etc. more efficiently. I am sure this conference will provide a platform to the young researchers to present their work and gain knowledge from others experience. It will provide an opportunity to collaborate and forge new alliances

I will like to congratulate Lt Gen P D Bhargava, the conference chair and Group Dy Vice Chancellor, Professor Raj Kamal Kapur, Dy. Dir & Head AITTM, convener TEL-NET2013 and his dedicated team for taking this initiative and working untiringly to ensure the success of their first Conference.

I wish the conference a grand success.

Mr Atul Chauhan
Chancellor
Amity University
Uttar Pradesh, NOIDA

Message from Group Vice Chancellor Amity Universities



I am happy to know that Amity Institute of Telecom Technology and Management (AITTM) and IETE Sub-Centre Amity University, NOIDA are organizing the first International Conference on Telecommunication and Networks (TEL-NET 2013).

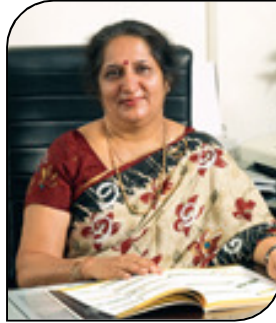
This is the first International Conference being organised on the theme of Telecommunication and Networks, both of which have become inseparable part of our daily life. Today, Telecommunication has brought revolution in the way we exchange information. Knowledge is power; and telecommunication and Information technologies are fast transforming our society to a knowledge society. It has completely changed the way we conduct our business. This conference, I am sure will further give inputs in research and innovation in these two important fields.

I congratulate Lt Gen P D Bhargava, the Conference Chair and Group Dy. Vice Chancellor & Advisor – AITTM, and Professor Raj Kamal Kapur, Convenor Tel - Net 2013, Dy. Director & Head – AITTM and his dedicated team for taking this initiative and working untiringly to ensure the success of this International Conference.

I also extend a warm welcome to all the participants and speakers to this Conference and I am sure that exchange of views and knowledge of eminent experts and professionals in this area would bring tangible outcome and the way forward in the future.

Maj Gen K Jai Singh
Group Vice Chancellor
Amity Universities &
Executive Senior Vice President, RBEF

Message from Acting Vice Chancellor



It gives me a great pleasure to know that Amity Institute of Telecom Technology and Management (AITTM) and IETE Sub-Centre Amity University, Noida are organizing its first International Conference on **Telecommunication and Networks (TEL-NET2013)** on 27-28 February 2013.

Telecommunication along with the Information Technologies is the most important drivers for socio economic development of a nation. We, as a nation, are reaping the benefits of transforming our society into a knowledge based society but we are still lagging behind in the research and technology in this field. I am sure, holding a Conference on Telecommunication and Networks is a right step to initiate research and collaboration between Academia, Scientific Community and Industry. The Conference is focusing on the current and future technologies in the field of telecommunications. The deliberations during the conference will offer new challenges and opportunities for the researchers and all other participants. I am sure it will widen our knowledge and will give direction to meaningful research in this field.

I welcome all our esteemed speakers and participants for enriching deliberations during the conference.

I also take this opportunity to congratulate Lt. Gen. P D Bhargava, Group Dy. Vice Chancellor & Advisor, AITTM for his initiative. I am confident that Prof. R.K. Kapur, Convenor **TEL-NET 2013**, Dy. Director & Head, AITTM and his team will do their utmost to come up to the expectations of all.

I wish the Conference all the success.

Professor (Dr.) Balvinder Shukla
Vice Patron TEL-NET 2013
Acting Vice Chancellor
Amity University Uttar Pradesh

Message from Group Deputy Vice Chancellor Amity Universities



It is a great honour for Amity Institute of Telecom Technology and Management (AITTM) and IETE Sub-Centre Amity University UP, NOIDA to organize the first International Conference on Telecommunication and Networks (TEL-NET 2013) on 27 and 28 February 2013 in association with The Institution of Electronics and Telecommunication Engineers (IETE).

The Conference was conceived with the idea to give fillip to the research activities in the field telecommunication and networking, expose our students to latest advancements in this field and to forge collaboration between academia, Industry and other scientific community. The response to the first such Conference has been very encouraging and it has strengthened our belief that such initiatives will benefit all participants.

Telecommunications is a vital support infrastructure for growth and modernisation of different sectors of economy. The unprecedented increase in the tele-density in the last decade and half has contributed substantial growth in the GDP of our nation. The number of users and the complexities of the networks are increasing exponentially, these are posing challenges for development of new technologies to meet the user aspirations. The biggest challenge however remains in the development of human resource with skill set to operate the complex systems and undertake the research and development in this field. I am sure this Conference will work out a road-map to meet some of the challenges.

I welcome all the participants and speakers to the conference and hope they benefit as much as we do during the Conference. I will also like to compliment Professor Raj Kamal Kapur, Dy. Dir & Head AITTM, convener TEL-NET2013 and his dedicated team for their efforts in ensuring the success of the Conference.

I wish the Conference all the success.

Lt Gen P D Bhargava
Conference Chair
Group Deputy Vice Chancellor
Amity Universities

Contents

Preface	i
Acknowledgement	ii
Committees	iii
Organizing Committee	iv
Speakers	v
National Advisory Committee Members	vi
International Advisory Committee Members	vii
Founder President's Message	viii
Message from Chancellor Amity University Uttar Pradesh	ix
Message from Group Vice Chancellor Amity Universities	x
Message from Acting Vice Chancellor	xi
Message from Group Deputy Vice Chancellor Amity Universities	xii
1. Free Space Optical Communication: Laser Sources, Modulation Schemes and Detection Techniques	1
<i>Anshul Vats, Hemani Kaushal, V.K. Jain</i>	
2. Model and Simulation of Different pumping Methods in Erbium- Doped Fiber Amplifier	6
<i>Md Mustafa Kamal</i>	
3. Mode Delay of Optical Waveguide Using Equivalent TL Circuit	10
<i>Vinita Mathur, Parul Tyagi, Surbhi Jain</i>	
4. Metamaterials: A New Boon in Design of Frequency Selective Surfaces	13
<i>Reepika Sharma, Ankush Kapoor</i>	
5. Performance Analysis of OCDMA System using Various Amplifiers	18
<i>Charu Singh, Brahmraj Singh, Brijesh Jaiswal</i>	
6. Review of ICI (Intercarrier Interference) Cancellation Techniques used in OFDM System	22
<i>Ritika, Gurpriya Sandhu, Garima Saini</i>	
7. A Survey on Fiber Optic Sensors	26
<i>Pinky Khundrakpam, Priyanka Sharma</i>	
8. Holography: Real Cyber World	29
<i>Aprajita Sharma, Ashish Gupta</i>	
9. Performance Evaluation of DWDM Systems by Using Various Amplifiers	33
<i>Amit Kumar Gautam, Brahm Raj Singh, Gurjit Kaur</i>	
10. De-De Dodging Algorithm for Scheduling Multiple Instances of Multiple Workflows in Hybrid Cloud	37
<i>Arun Kumar. B, Ravichandran. T, Sundareswari. K</i>	
11. An Investigation of Cloud Computing	42
<i>Shubhani, Anil Kumar Gankotiya</i>	
12. Analysis of MEMS Application	48
<i>Megha Goyal, Dolly Gupta</i>	
13. 4G: A New Regime in the Mobile Communications Generations	52
<i>Rekha Kashyap, Ankush Kapoor</i>	
14. 5G Technologies	56
<i>Vinay Kumar Singh, Shekhar Singh, Rachit Manchanda, Shilpa Thakur</i>	

15.	Role of Telecommunications Network in Universities	59
	<i>C.G. Nayak, Balbir Singh</i>	
16.	A New More Efficient, Dynamic, and Robust Access Control Scheme over Wireless Sensor Networks	63
	<i>Raj Kumar, Ritesh Kumar</i>	
17.	Mobility Management Issues in Hierarchical Mobile IPv6 in 4G Networks	70
	<i>Shamurailatpam Susanta Sharma, Himanshu Sharma</i>	
18.	Use of Storage as Service for Online Operating System in Cloud Computing	76
	<i>Piyush Saxena, Satyajit Padhy, Praveen Kumar</i>	
19.	Overview of Motion Estimation Algorithms in Video Compression for H.261 ITU-I Recommendation	82
	<i>Rahul Bhandari, Eshank Jain</i>	
20.	Investigation of XGM effect of SOA for OOK, DPSK, Duo-binary & Manchester format	87
	<i>Manish Chauhan, Hemant Purohit</i>	
21.	Retrieval of Target Velocity using Doppler's Effect Phenomena by Comparative Study of L, C and X Band Radar	91
	<i>Garima Sharma, Nishank Agarwal, Aditya Sharma, Sudhir Kumar Chaturvedi, Pavan Kumar Nanduri, Ugur Guven</i>	
22.	Enhanced QoS Through Traffic Pacified Handoff Algorithm in Wimax 16m Networks	94
	<i>D. Karunkuzhali, D.C.Tomar</i>	
23.	Enhancing Interaction Study on the Cloud Computing	102
	<i>Mohit Bhansali, Praveen Kumar, Seema Rawat</i>	
24.	Overview of IPv6 and its Implementation	109
	<i>Nistha Rai, Kanak Priya, Neha Kumari</i>	
25.	Synchronization Techniques for OFDM Systems	112
	<i>Pratima Manhas, Shaveta Thakral</i>	
26.	A Review Paper on Zigbee	117
	<i>Amit Verma, Anvita Tripathi, Kapil Kumar</i>	
27.	New Emerging Era of Communication: Broadband Technology - A Review	124
	<i>Mitesh Sharma</i>	
28.	Performance Analysis with Multi-antenna for MIMO Wireless System	130
	<i>Monalisa Bhol, Sikha Mishra, Mihir Narayan Mohanty</i>	
29.	Industrial Ethernet: Convergence of Communication and Information Technology with Control Systems	134
	<i>H.K. Gopinath, Ajeet S Rawat, Mukesh G, Rakshit G.</i>	
30.	Security Management on Telecommunication Network	140
	<i>Abha Jaiswal, Ravi Prakash Jaiswal</i>	
31.	An Artificial Neural Network Based Approach of Rejecting Static Clutter from the Covariance Matrix Containing Low Velocity, Small RCS Target in Case of Monostatic Airborne Radar	144
	<i>M Chakraborty, P Karmakar, B. Maji, D. Kandar</i>	
32.	Plug and Play VOIP with Raspberry Pi for Small Scale Industries	148
	<i>Chirag Gohel, Himanshu Madhavani</i>	

33.	VoIP over Office Network	154
	<i>Sushil Kumar</i>	
34.	An Analysis of Unicast Routing Protocols in Mobile Adhoc Networks (MANETS)	161
	<i>Inhas Ashraf, Shabir A.Sofi, Sheikh Obaid Ahmad, Bilal Ahmad Yatoo</i>	
35.	Review on Recent Energy Efficient Techniques in Wireless Sensor Networks	168
	<i>Jyoti, Ranjana Thalore, Manju, Urvashi Singh, M. K. Jha</i>	
36.	Study of Proactive Routing Protocol for different Buffer Size	173
	<i>Yashi Rajvanshi, Seema Rahul, Sanjay Maurya, Mayank</i>	
37.	Application Layer Multicast Protocols	177
	<i>Shubha Shukla, Akhilesh Kosta, Rohit Kumar</i>	
38.	Analysis of Energy Consumption, Throughput and Delay using Modified EQSR Routing Protocol in Wireless Sensor Network	184
	<i>Sunita, O.S. Khanna, Amandeep Kaur</i>	
39.	Data Dissemination Protocol Based on Different Groups of Grid in WSN	191
	<i>Divya Sharma, Chandni, Kanika Sharma</i>	
40.	Cluster-Based Routing Protocols for Heterogeneous Wireless Sensor Networks	196
	<i>Suniti Dutt, O. S. Khanna</i>	
41.	Security Issues in Ad Hoc Networks: A Survey	201
	<i>Rumisa Firdous, Emmanuel S. Pilli, Shabir Ahmad Sofi</i>	
42.	Review of Chain Based Hierarchical Protocols in Wireless Sensor Network	206
	<i>Richa Mehta, Sandeep Verma, O.S. Khanna</i>	
43.	An Adaptive Cross Layer Routing Mechanism to Optimize Qos in MANET	212
	<i>V. Dhilip Kumar, D. Kandar, C.K. Sarkar</i>	
44.	A Study on Future Advancements in Security for MANET	216
	<i>Samta Suman Lodhi, Radhey Shyam Lodhi</i>	
45.	A Comparative Study of Security Attacks in Bluetooth, Wi-Fi and Wimax	220
	<i>Nandini Deb, Tushar Saxena, Himanshu Goyal</i>	
46.	Study on Combining 3G, Wi-Fi and Wi-max for Wireless Broadband	227
	<i>Nistha Rai, Kanak Priya, Neha Kumari</i>	
47.	A Comparison Study of Dynamic Addressing Techniques in Mobile Ad hoc Networks	233
	<i>Navin Paudel</i>	
48.	Performance Metrics for Proactive and Reactive Routing Protocols in Mobile Adhoc Network	239
	<i>Basu Dev Shivhare, Anil Kumar Sajnani, Shalini Shivhare</i>	
49.	Wireless Sensor Networks with application to the Measurement and Detection of Air Pollution	244
	<i>Apurv Gupta, Rohit Kathait, Ankush Kapoor</i>	
50.	Routing of Autonomous Wheeled Mobile Robot in External Environment using Wireless Sensor Network	249
	<i>Anuj Chadha</i>	
51.	A Discussion on Hardware Platforms for Wireless Sensor Network	253
	<i>Malang Shah, Saurabh Mehta</i>	

52.	An Overview of Localization Techniques and Algorithms for Wireless Sensor Network	261
	<i>Sudhir P. Kasar, Ashish K. Shekhar, Malang Shah, Saurabh Mehta</i>	
53.	Securing Mobile Phone Communications	266
	<i>Abubaker Maraicar</i>	
54.	Security Challenges in Cloud Computing and SAML: A Study	275
	<i>Sarah J. Andrabi, Inhas Ashraf, Roohie Naaz Mir, Shabir A. Sofi</i>	
55.	Implementation of Digital Signature Algorithm for Improved Performance for Small Data Sets	281
	<i>Vijay Kumar Tiwari, Anuraag Awasthi, Ritesh Rastogi, Anuj Kumar</i>	
56.	A Semantic Exert to Implement a Novel Framework for E-Learning Web Using Protégé	287
	<i>Akhilesh Dwivedi, Aparna Bawankan</i>	
57.	Comparison between Agile and Traditional Software Development Methodologies & Design A Hybrid Software Development Methodology	293
	<i>Iti Kapoor, Prem Sagar Sharma, Atul Kumar Srivartava</i>	
58.	Unique Watermark Generation Using LFSR and EBCDIC Code	299
	<i>Shaikh Rakhshan Anjum, Priyanka Verma, Asna Furquan</i>	
59.	Enterprise Portal with Java Portlet for Universities	302
	<i>Veenita Gupta, Neeraj Kumar, Seema Rawat, Praveen Kumar</i>	
60.	Attacks and its Security Mechanism in AODV for Mobile Ad hoc Networks	307
	<i>Mahak Gupta, Kushagra Agrawal, Rajneesh Kumar Gujral, Sanjeev Rana</i>	
61.	Network Layer Attacks and its Countermeasure in Mobile Ad hoc Networks	314
	<i>Avinka Baweja, Kushagra Agrawal, Sanjeev Rana, Rajneesh Kumar Gujral</i>	
62.	Permission-Based Security Models and its Application to Android System	321
	<i>Ahmad Talha Siddiqui, Munesh Chandra Trivedi</i>	
63.	Computational Intelligence in Wireless Sensors Networks	326
	<i>Janani Rajaraman</i>	
64.	Wireless Telecommunication Technologies Vulnerability and Attacks on Wireless Systems	334
	<i>Ankur Agrawal, Utkarsh Sharma</i>	
65.	Software as a Service (SaaS) Approach for E-learning	338
	<i>M. K. Sharma, Manisha Gururani</i>	
66.	Survey on Security Issues in Cloud Environment	342
	<i>Sudhir Shenai, M. Aramudhan, A. Devi Priya</i>	
67.	Single Document Summarization Using TF/IDF Technique	348
	<i>Nandini Anand, Mayank Baheti, Prakhar Rastogi, Atul Kumar Srivastava</i>	
68.	Language Morphology and Search Engine Performance	354
	<i>Nargis Parveen, Mohd Athar</i>	
69.	Optimization of Software using Genetic Algorithm and Formulation of Test Suite for Code Coverage	361
	<i>Mohd Athar, Avdhesh Gupta</i>	
70.	A Study on Digital Image Watermarking	368
	<i>Himanshu Goyal, Rahul Raj, Arpan Batra, Kaushlendra Singh</i>	

71.	Digital Image Watermarking Algorithm Based on DCT and Spread Spectrum	372
	<i>Harsh Vikram Singh</i>	
72.	Koch Shaped Fractal End Coupled Microstrip Bandpass Filter	378
	<i>Gurpreet Kaur Kohli, Pravesh Singh</i>	
73.	New Fast and Efficient Progressive Switching Median Filter for Digital Images	381
	<i>Ritesh Kumar, Lav Kedia</i>	
74.	A Survey on Capacity Analysis for Multiuser MIMO Downlink System in Wireless Communication	387
	<i>Abhishek Gupta, Garima Saini</i>	
75.	Channel Performance by using Adaptive Equalization Techniques in MIMO System for Multipath Fading Environment	392
	<i>Geetesh Kwatra, Liladhar Malviya</i>	
76.	Medical Image Retrieval Using Texture Features	397
	<i>A. Swarnambiga, S. Vasuki, A. Anantha Raja</i>	
77.	Reconfigurability in Microstrip Patch Antennas	402
	<i>T. Sushma, N.V. Koteswara Rao, K. Rama Naidu</i>	
78.	Medical Image Registration Based Retrieval Using Color and Texture Features	408
	<i>A. Swarnambiga, S. Vasuki, A. Ganesh Lakshmanan</i>	
79.	A Multi Band Switchable Circularly Polarized Slotted Microstrip Patch Antenna	415
	<i>Ajit Yadav, Shweta Gautam, Mithilesh Kumar</i>	
80.	A Reconfigurable Multiband Square Patch Antenna	419
	<i>Shweta Gautam, Ajit Yadav, Mithilesh Kumar</i>	
81.	Design of Compact BPF for UWB Communication using Multi-Physics	423
	<i>Malabi Singh, Mihir Narayan Mohanty</i>	
82.	Target Position Estimation by Synthetic Aperture Radar (SAR) Dataset	427
	<i>Ganesh Dutt, Kolli Sridatta Sairam Reddy, Ishan Sharma, Sudhir Kumar Chaturvedi, Ugur Guven, Pavan Kumar Nanduri</i>	
83.	Anthropogenic Treaty	430
	<i>Shefali Nagar, Shikha Puri, Priyanshi Dwivedi</i>	
84.	A Novel Geometry of Wideband Microstrip Patch Antenna with Finite Ground Plane	433
	<i>Sanyog Rawat, K. K. Sharma</i>	
85.	Stacked Multiband Triangular Fractal Antenna for Mobile Communications	438
	<i>Sumit Kumar, Richa Sharma</i>	
86.	An Enhancement in Data Compression Using H.264 /AVC	442
	<i>Ankita Awasthi, Anshika Salaria, Samta Suman Lodhi</i>	
87.	Study of Smart Antennas and their use in Wireless Communication Systems	445
	<i>Amritpal Singh Bhinder, Rajat Singh</i>	
88.	Parametric Analysis of Co-axial Probe fed Rectangular Dielectric Resonator Antenna	448
	<i>Neeraj Kumar, Arvind Kumar</i>	
89.	Motion Detection and Tracking of Video Sequences: A Survey	451
	<i>Neha Kumari</i>	

90.	13T Low Power PTL based Arithmetic Leaf Cell for Signal Processing	457
	<i>S.Vijayakumar, Reeba Korah</i>	
91.	Large Scale Path Loss Outdoor Propagation Models: A Survey	465
	<i>Richa Budhiraja</i>	
92.	An Overview of Neural Filters for Impulse Noise Removal	469
	<i>Rashmi Kumari, S.K. Aggarwal</i>	
93.	MOSFET's The New Generation Transistors	474
	<i>Rahul Gautam, Ankush Kapoor, Himanshu Saxena</i>	
94.	Integrated Circuits and MicroElectroMechanical Systems Fabrication A Review on Pre-existing Fabrication Techniques	480
	<i>Himanshu Saxena, Rahul Gautam, Ankush Kapoor</i>	
95.	Comparison between the Implementation of data Compression using Huffman Coding and Shannon- Fano Coding through VHDL	487
	<i>Pooja Srivastava, Jyotsna Joshi, Abhishek</i>	
96.	Design and Performance Analysis of Sound Level Meter	493
	<i>Sushil Kumar</i>	
97.	Microcontroller Based Data Analyzer through Wireless RF	498
	<i>Akshay Nangia, Madhurima Basak, Nippun Bahl</i>	
98.	Design and Performance of Eleven Stages CMOS Ring Oscillator in 45 nm Technology	501
	<i>Sushil Kumar</i>	
99.	Microcontroller Based Safety Guard System for Blind	506
	<i>Ruchi Gupta, Alka Verma, Okar Singh, Renu Pooniya</i>	
100.	Heuristic Algorithm Approach for FSM Decomposition to Reduce Power	510
	<i>Himani Mittal, Naved, Predeep Kumar</i>	
101.	Independent Control of Rise and Fall Edge Dead Time Using IC 555 Timer	514
	<i>Ritish Kumar, Jitender Kumar</i>	
102.	Spatial Filters: A FSS Approach of Designing	516
	<i>Vishakha Kanwar, Ankush Kapoor</i>	

Free Space Optical Communication: Laser Sources, Modulation Schemes and Detection Techniques

Anshul Vats¹, Hemani Kaushal², V.K. Jain³

^{1,2}Department of EE&CE, ITM University, Gurgaon-122017 (Haryana), India

¹anshul11ecp005@itmindia.edu, ²hemanikaushal@itmindia.edu

³Department of Electrical Engineering, Indian Institute of Technology,
New Delhi-110016, India, vkjain@ee.iitd.ac.in

Abstract: Free space optical (FSO) communication is an upgraded supplement to existing wireless technologies. FSO technology has less power and low mass requirement, vast modulation bandwidth, unlicensed spectrum and cost effective deployment. FSO systems are used to transmit and receive all kind of data at high data rates (up to 80 Gbps). Today researchers are preliminary focused to use the free space optical communication systems for designing ground to satellite inter- satellite and satellite to ground optical link. This paper gives a review on the laser diodes, modulation schemes and the detection techniques which are deployed in the FSO communication systems.

Keywords: Free space communication, Modulation schemes, Laser diodes, Detectors.

1. INTRODUCTION

Laser communication or free space optical communication is a technology that involves transmission of information on to the optical carrier through the space from source to destination. FSO system uses lasers to produce optical signals of narrow beam divergence this narrow beam focuses a large amount of transmitted signal power on to the receiver and hence gives the higher link power efficiency. FSO systems enable 10 to 100 times more data transmission by utilizing only 1% of antenna aperture area when compared with RF antennas of wireless communication [13]. It utilizes less power and mass, provides a secure and jam free network and has unlimited bandwidth with no regulation on the optical spectrum.

These systems have wide range of applications- connecting sites in an area, extending the fibre optic cable network to the nearby buildings, in local loop bypass, backhaul, disaster recovery, in last mile applications, inter satellite links, links between spacecraft and satellite and many more. Figure 1 shows the application scenario of FSO technology. These systems operate very much like a fibre optic connection which uses a fibre. The main difference is that the attenuation from the cable is known and can be controlled. But the FSO link uses space/atmosphere as the media and the attenuation may vary every second and is unknown.

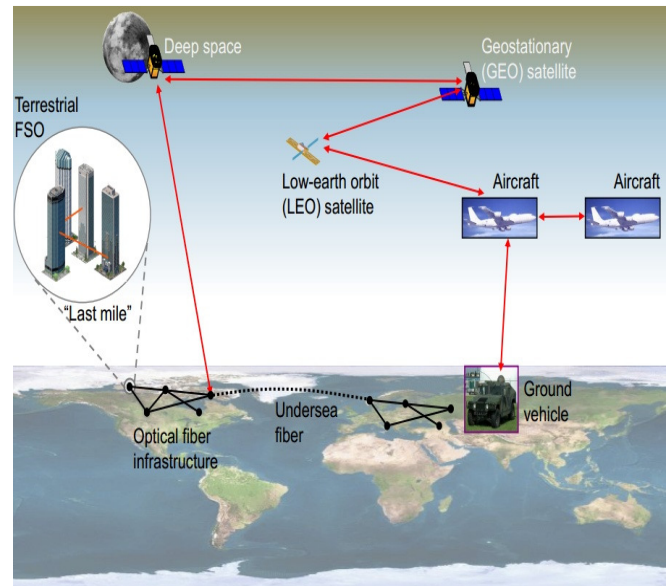


Fig. 1. Free space optical communication applications

FSO communication system uses laser diodes whose operating wavelengths are ranging from 780 nm to 900 nm and 1500 nm to 1600 nm. Figure 2 gives the block diagram of FSO communication system. In this block diagram, laser diodes are used at the transmitter end to produce narrow beam optical signals. These optical signals are modulated by information signal and then transmitted towards the receiver. The laser diode produces a narrow beam on which information signal is modulated and then transmitted towards receiver [2]. Various modulation schemes used in FSO communication are On-Off Keying (OOK), Pulse Position Modulation (M -PPM) and Subcarrier Intensity Modulation (SIM). Now a days, Differential Phase Shift Keying (DPSK) is also getting popular. The received optical signal is then detected by PIN or Avalanche Photo Diode (APD) which convert the optical signal to electrical signal for further processing. The signal from the detector is then given to demodulators to extract the original information [4]. The laser diodes, modulation schemes and detection techniques are discussed in the subsequent sections.

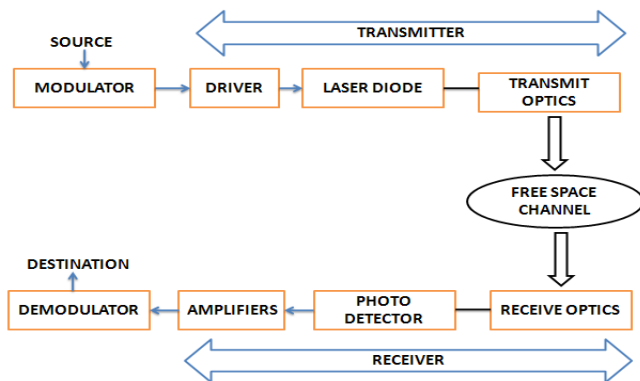


Fig. 2. Block diagram of the FSO link

2. LASER DIODES

For FSO communication, laser diodes are preferred over LEDs, because their beam does not spread while covering longer distances. Also they have higher frequencies which increase the modulation rate and overall communication rates. They have faster rise and fall times which enhances the switching speed and over all throughput of the system.

Those laser diodes which have operating wavelength centred at 800 nm and 1550 nm are generally preferable for FSO systems and Inter-Satellite Link. We can design an eye safe laser transmitter in these transmission windows. Between them 1550 nm wavelength is preferred because allowable safe laser is fifty times more. This gives up to 17 db extra margin and make system to propagate over longer distances with higher data rates. Among solid state laser, Nd: YAG (neodymium yttrium aluminium garnet) which operates at 1064 nm wavelength is most widely used. This laser is adequate to transmit immense amount of power (20- 30 W) and is used in coherent systems with highly stable Nd: YAG oscillator [18]. Table I shows the laser technologies which are commonly used for FSO systems [3] and Table II gives the compounds involved in the lasers diodes [11].

Table I. Laser Sources [3]

S. No.	Technology type	Operating wavelength	Features
1.	Vertical Cavity Surface Emitting Laser	~ 850 nm	Lower power density Cheap and readily available. No active cooling High reliability Output optical power: max up to 20 mW and typical power: 6 mW. Low threshold and operating current. 8.5 Gbps data rate and reliable up to 10 Gbps. Applications are- optical fibre communications, computer mice, gas sensing, optical clocks.

2.	Fabry Perot	1300/1550 nm	50 times higher power density. Long life Low eye safety criteria. Output Optical power: 20 mW-100 mW and typical power: 28 mW. ±0.03 db CW power stability. Insensitive to back reflection & stabilised for short & long term application. Narrow spectral resolution. Internal digital modulation. Upto 40 Gbps data rate Applications are- in dichroic filters, add-drop multiplexers with banks of miniature tuned fused silica or diamonds, optical wavemeter, laser resonator, laser absorption spectrometry techniques, in gravitational wave detection.
3.	Distributed Feedback Lasers	1300/1550 nm	Compatible with EFDA Higher data rates upto 40 Gbps. Small temperature dependence. Complex fabrication Narrow emission linewidth of < 1 nm. Provide superior longitudinal mode discrimination over Fabry perot. Output optical power: >20 mW and typical power: 1-2 W when combined with EFDA. ± 0.01 db CW power stability Applications are- DWDM, CATV and long haul communication.
4.	Solid State Lasers	1064 nm	High power in infrared spectrum. Small gain bandwidth of the order of 1 nm or less. Very good coherence and suitable for homodyne systems. Natural birefringent. Laser gain is strongly polarization dependent. Applications are-ophthalmology to correct posterior capsular opacification, flow visualization techniques in fluid dynamics, soft tissue surgeries , laser designators and laser rangefinders, cavity ring-down spectroscopy, laser pumping, laser induced break-down spectroscopy.

Table II. Compounds Used In Laser Diodes [11]

S. No.	Operating Wavelength	Compound(s)
1.	620-895 nm	Ga _(1-x) Al _(x) As
2.	904 nm	GaAs
3.	1100-1650 nm	In _(1-x) Ga _(x) As _(y) P _(1-y)
4.	1550 nm	In _(0.58) Ga _(0.42) As _(0.9) P _(0.1)
5.	1604 nm	Nd ³⁺ :Y ₃ Al ₅ O ₁₂ ; Nd ³⁺ :YVO ₄ ; Nd ³⁺ :YLiF ₄

3. MODULATION SCHEMES

Different modulation schemes exist which are well suited for the free space optical communication. Most commonly used modulation schemes are On-Off Keying (OOK), Pulse Position Modulation (PPM), Differential Phase Shift Keying (DPSK), Differential Quadrature Pulse Shift Keying (DQPSK) and Subcarrier Intensity Modulation (SIM). OOK is the simplest modulation scheme (because of its design and implementation) in which transmitter is ON only for the binary bit '1' and OFF for bit '0'. OOK modulation scheme is widely commercially available for the applications of FSO systems [12]. Though it is simple to implement, it is not an optimal modulation scheme. As the atmospheric channel is a continuous varying channel, it requires adaptive threshold under varying turbulent conditions [11].

Another modulation scheme is M -ary Pulse Position Modulation (PPM). It is well suited for direct detection of optical signal transmitted through wireless space. PPM offers a great advantage of eliminating the decision threshold dependencies on the input power unlike OOK modulation. Therefore it is a power efficient modulation. The main limitation is that it requires more bandwidth than that of OOK [6]. PPM needs a complex transmitter and receiver designs circuits because of high level of synchronisation needed between different time slots. Increase in value of M leads to increase in transmission efficiency but at the cost of increase in required bandwidth by the factor of $M/\log_2 M$. Therefore huge value of M reduces the band utilization efficiency. In M -PPM more difficulties occurred in the recovering the symbol timing reference [6].

Another type of modulation popularly used these days is DPSK. This scheme encodes the data bits on its phase, can extenuate serious effects of scintillation to some extent. DPSK has benefit over OOK, that it has ~3db lower optical SNR needed to obtain a given BER if a balanced receiver is used [6]. For OOK quantum limit for an optically pre-amplified receiver is 41 photons/bit, this reduces to 22 photons/bit with a balanced detector. It gives higher data rates over PPM and OOK with increase in complexity in receiver. In DPSK, bandwidth decreases linearly with decrease in data rates thus it is not suitable for the lower data rates. Also, its receiver requires single mode optical signal free from phase noise which decreases the collection efficiency of the signal. Due to all these limitations, the use of DPSK modulation scheme in the turbulence free links such as between satellites or ground to satellite is limited [6].

When we compare the binary modulation schemes like OOK, DPSK with the DQPSK scheme, it doubles the spectral efficiency by making advantage of two signal quadrature of an optical carrier signal [5, 6].

Subcarrier Intensity Modulation (SIM) is again one of FSO modulation scheme. It does not require the adaptive threshold (like in OOK) and not need much bandwidth (like in PPM). SIM has a drawback that it suffers from a high peak to average power ratio, thus giving poor power efficiency. Also, the non-linearity of the component is a big issue when dealing with multiple subcarriers. One has to choose the modulation schemes as per need of the application with some trade off among described factors.

4. OPTICAL DETECTION TECHNIQUES

This section of paper tells about the photo detection techniques and photo detectors which convert the received optical signal into corresponding electrical signal for the further signal processing or decision making at the receiver. Photo detectors primarily extract the information embedded on the optical carrier signal (it may be embedded on frequency, phase or intensity of the optical signal).

Photo detectors are the transducers which convert optical signal to the corresponding electrical signal. They should have high sensitivity within its operational wavelengths, low noise levels and has sufficient bandwidth to hold the needed data rates. Detectors should have minimum effect on the response of detector due to temperature fluctuations. Device should have long operating life too [21].

Two types of photo detectors are used in FSO-PIN and Avalanche Photo Diode (APD). PIN photo diodes have a P and N type semiconductor layer separated by a very lightly n type doped intrinsic layer [15, 21]. The responsivity of PIN photodiode is always less than unity. APD photodiode provides an inherent current gain which increases the sensitivity of the detector. Typical values for APD gain are in the range of 50-300 [17, 21]. This implies responsivity of APD is greater than unity. APD provides the higher sensitivity as compared to PIN detector but the statistical nature of the ionization/avalanche process means that there is always a multiplication noise associated with the APD. The avalanche process is also very temperature-sensitive, but has multiplication noise and is very much temperature sensitive too [9, 21].

Table III. Photo detector's material and corresponding wavelength and energy gap [15, 21]

S. No.	Material	Wavelength (nm)	Energy gap (eV)
1.	InGaAsP	1650-920	0.75-1.35
2.	InGaAs	1700	0.73
3.	GaAs	870	1.424
4.	Germanium	1600	0.775
5.	Silicon	1060	1.17

At the receiver, depending upon whether the local oscillator is used or not in the detection, following detection techniques can be used.

- Direct Detection
- Coherent Detection
- Heterodyne Detection
- Homodyne Detection
- Maximum likelihood sequence Detection (MLSD)
- Iterative Detection

A. Direct Detection

In direct detection information is encoded with the intensity variations. Here there is no need of local oscillator for detection. Hence no synchronisation is needed between receiver and transmitter. This is also called envelope detection [17, 21].

B. Coherent Detection

In coherent detection, local oscillator is needed to obtain optical signal operating at a particular wavelength. The frequency and phase of local oscillator need not to be same as that of received signal [19, 21].

1. Heterodyne Detection

In heterodyne detection, the frequency and phase of local oscillator need not to be same as that of received signal. The received signal is mixed with a reference wave from local oscillator on the photo detector. It is relatively easy way of amplifying the photo current by increasing the local oscillator power. This detection provides the improved SNR by increasing the local oscillator power. IF frequency needed to monitor regularly so as to maintain the IF centre frequency constant. Noise is also another limitation which is contributed by shot noise, photo detector's noise and added by electronics. These are factors which are the challenges while implementing the coherent optical communication system [19, 21].

2. Homodyne Detection

Homodyne detection is similar to the heterodyne detection technique. But here local oscillator has the same phase and frequency as that of the received optical signal [19, 21].

C. Maximum Likelihood Sequence Detection (MLSD)

Optimum multi-user MLSD applies maximum likelihood principle- Considering the whole received sequence; find the estimate for the received sequence that has the minimum distance to the allowed sequences. It has the optimum

performance provided transmitted symbols equal alike has large computational complexity. In exhaustive search 2^{NK} vectors to be considered (K users, N bits) requires estimation of received amplitudes and phases that takes still more computational power. It can be implemented by using Viterbi-decoder that is 'practically optimum' ML-detection scheme to reduce computational complexity by surviving path selections.

D. Iterative Detection

The iterative detection and decoding is performed by iteratively passing soft (multi-bit) "a priori" information between a detector and a decoder. The detector receives modulation symbols, performs a detection function that is complementary to the symbol mapping performed at the transmitter, and provides soft-decision symbols for transmitted coded bits. "Extrinsic information" in the soft-decision symbols is then decoded by the decoder to provide its extrinsic information, which comprises the "a priori" information used by the detector in the detection process. The detection and decoding may be iterated a number of times.

5. CONCLUSION

This paper gives a brief review on various optical laser diodes which can be utilized in the free space communication system. These laser diodes can operate on different IR spectrum windows such as 850 nm, 1064 nm and 1550 nm to give higher data rates and provide spectrum which is unlicensed and without any regulations on it. So we can freely use the band with higher capacities. They can be used in various applications of FSO communication such as inter satellite links, air borne to satellite link, last mile solution, LAN, MAN and many more. Besides laser diodes modulation schemes are also discussed which provide a brief idea about modulation schemes suited for a particular FSO application. Finally detectors and detection techniques are briefed which are used in FSO application.

REFERENCES

- [1] N. Kumar, V.K. Jain, S. Kar, "Evaluation of the performance of FSO system using OOK and M -PPM modulation schemes in intersatellite links with turbo codes," in *International conference on electronics, communication and computing technologies (ICECCT '11)*, Paul Nagar, pp. 59-63, Sep 2011.
- [2] G. Soni and J. Malhotra, "Free space optics system: performance and link Availability," *International journal of computing and corporate research*, vol 1, no. 3, Nov 2011.
- [3] Z. Ghassemlooy and W.O Popoola, *Mobile and wireless communications, Mobile and Wireless Communications Network layer and circuit level design*, Publisher: INTECH, 1ed. Jan 2010.
- [4] J. D. Moores, G. Frederick Walther, et.al, "Architecture Overview and Data Summary of a 5.4 km Free-Space Laser

- Communications Experiment,” Proc. of SPIE vol. 7464 pp.746404-1, Mar 2010.
- [5] Z. Wang W. D. Zhong, S. Fu C. Lin, “Performance Comparison of Different Modulation Formats Over Free-Space Optical (FSO) Turbulence Links With Space Diversity Reception Technique,” *IEEE Photonics Journal*, vol 1, pp.277-285, Dec 2009.
- [6] N. Chand, A. J. Hunton, and B. M. Eteson, “A comparative study of 2.667 Gb/s OOK, DPSK, and PPM modulation formats for FSO applications,” *SPIE Free space laser communication VIII*, vol. 7091, pp. 20-27, Jan 2008.
- [7] W. O. Popoola, Z. Ghassemlooy, and E. Leitgeb, “Free-space optical communication using subcarrier modulation in gamma-gamma atmospheric turbulence,” *9th International Conference on Transparent Optical Networks (ICTON '07)* vol. 3, pp. 156-160, Jul 2007.
- [8] H. Hemmati, “Deep Space Optical Communications,” in Deep space communications and navigation series Canifornia: Wiley-Interscience, Dec 2006.
- [9] K. Kiasaleh, “Performance of APD-based, PPM free-space optical communication systems in atmospheric turbulence,” *IEEE Transactions on Communications*, vol. 53, pp. 1455-1461, Sep 2005
- [10] S. Bloom, E. Korevaar, J. Schuster, and H. Willebrand, “Understanding the performance of free-space optics,” *Journal of Optical Networking*, vol. 2, pp. 178-200, Jun 2003.
- [11] M. Chen, Review of Free-Space Optical Communications Links, May 13th, 2003.
- [12] H. Willebrand and B. S. Ghuman, *Free Space Optics: Enabling Optical Connectivity in today's network*. Indianapolis: SAMS publishing, 2002.
- [13] The JPL website. [Online]. Available: <http://technology.jpl.nasa.gov/research/ResearchTopics/topicedetails/?ID=67>, last visit 9 Jul 2012.
- [14] H. Rongqing, Z. Benyuan, H. Renxiang, T. A. Christopher, R. D. Kenneth, and R. Douglas, “Subcarrier multiplexing for high-speed optical transmission,” *Journal of Lightwave Technology*, vol. 20, pp. 417-424, Jun 2002.
- [15] G. Keiser, *Optical Fiber Communications*, 3rd ed. New York: McGraw-Hill 2000.
- [16] S. Arnon Stanley, R. Rotman, N. S. Kopeika, “Performance limitations of free-space optical communication satellite networks due to vibrations: direct detection digital mode,” *Society of Photo-Optical Instrumentation Engineers*, vol. 36 no. 11, pp.3148–3157, Nov 1997.
- [17] R. M. Gagliardi and S. Karp, *Optical Communications*, 2nd ed. New York: John Wiley, 1995.
- [18] Sipes, D. L, Jr, “Highly efficient Nd: YAG lasers for free space communication,” The Telecommun. and Data Acquisition Rept, pp. 31-39 Jan 1985.
- [19] W. K. Pratt, *Laser Communication Systems*, 1st ed. New York: John Wiley & Sons, Inc., 1969.
- [20] R. Walter LEEB, “Space Laser Communications: Systems, Technologies, and Applications,” available: http://publik.tuwien.ac.at/files/pub-et_4235.pdf, last visited 20 jul 2012.
- [21] W. Popoola, “Subcarrier intensity modulated free-space optical communication systems,” University of Northumbria at Newcastle, Sep 2009.

Model and Simulation of Different pumping Methods in Erbium- Doped Fiber Amplifier

Md Mustafa Kamal

Department of Electrical Engineering
M.E Student National Institute of Technical Teacher Training and Research, Chandigarh
mustafakamalece@gmail.com

Abstract: Optical fiber system has distinct advantage over other transmission media for signal communication. some of the advantage of optical fiber system is wider bandwidth, better gain and low noise figure (NF). Because due to different transmission conditions and losses signal become weaker. So for amplification of signal different amplifiers are used. Erbium doped fiber amplifier (EDFA) is widely used amplifier for amplification. EDF amplifier has distinct advantage over conventional electrical amplifier. EDFA amplifier directly convert light to light signal with better amplification efficiency. For the amplification of signal erbium doped fiber amplifier need pumping signal. Depending upon the pumping methods amplification of different parameters (Gain, noise figure etc) of EDFA amplifier is different. In this paper we will discuss the different methods of pumping used in optical amplifier with the help of optical simulator optisystem 11.0 (license product of Canadian based company).

Keywords: Optical amplifier, optical pumping, forward pumping, backward pumping, bi-directional pumping, Gain, noise figure.

1. INTRODUCTION

In the optical fiber communication system there are various types of losses occurred during the transmission of signal. Some of these losses are attenuation losses, fiber joint losses; splices losses and fiber tap losses .Because due to above losses signal become weaker during the transmission. Hence signal which is transmitting from transmitter to receiver end get weaker. So we are not able to get appropriate strength of signal at the receiver end. For the amplification of signal different amplifier are used in the optical fiber system ,mainly used amplifier are optical based amplifier, semiconductor optical amplifier (SOA), fiber doped amplifier (EDFA and Raman amplifier).Erbium doped fiber amplifier (EDFA) has made tremendous progress since its invention in 1986.It totally replace the optical amplification technology in optical fiber system. Before using these amplifiers electrical amplifiers are used which convert light energy to electrical energy and then light energy by using other electrical amplifier, but by using EDF amplifier this is possible to directly convert the light signal to light signal. Hence EDFA amplifier not only

changes the involvement of repeater station in the optical medium but also this is very cost effective. It creates the revolution in long distance communication system [1].

2. STRUCTURE OF EDFA AMPLIFIER

Basically EDFA technology is the Erbium doped fiber (EDF), which is conventional silica fiber doped with erbium. When erbium is illuminated with light energy at suitable wavelength (either 980 nm or 1480nm)it is excited to a long lifetime intermediate state, following decay to ground state by emitting light within the band of 1525-1565nm.If light energy already exist within the 1525-1565nm band ,for example due to signal channel passing through EDF ,then this simulate the decay process (so called simulated emission), resulting in additional light energy[2] .Thus if a pump wavelength are simultaneously propagating through an erbium doped fiber, energy transfer will occur via the erbium from the pump wavelength to the signal wavelength ,resulting in signal amplification.

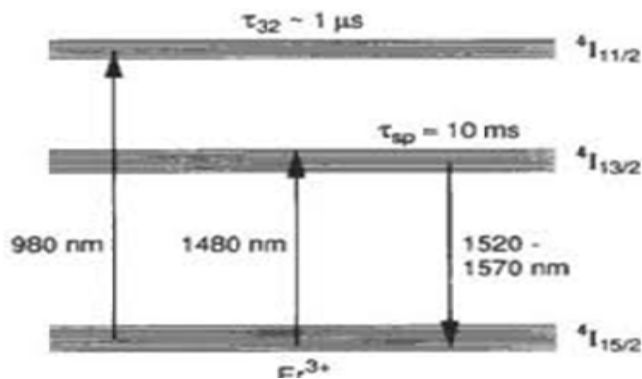


Fig. 1.1 Structure of EDFA energy diagram

3. PUMPING IN ERIBUM DOPEED FIBER AMPLIFIER (EDFA)

In its most basic form the EDFA consist of a length of EDF (typically 10-30m), a pump laser, and a component (often referred to as a WDM) for combining the signal and pump

wavelength so that they can propagate simultaneously through the EDF. In principle EDFA's can be designed such that pump energy propagates in the same direction as the signal (forward pumping), the opposite direction to the signal (backward pumping), or both direction together. The pump energy may either by 980nm pump energy, 1480nm pump energy, or a combination of both. Practically, the most common EDFA configuration is the forward pumping configuration using 980nm pump energy, as shown in Figure 2. This configuration makes the most efficient use of cost effective, reliable and low power consumption 980nm semiconductor pump laser diodes, thus providing the best overall design with respect to performance and cost trade-offs. Besides the three basic components described above, Figure 1.1 also shows additional optical and electronic components used in a basic single stage EDFA. The signal enters the amplifier through the input port, and then passes through a tap which is used to divert a small percentage of the signal power (typically 1-2%) to an input detector. The signal then passes through an isolator, before being combined with pump energy emitted by the 980nm pump laser diode[3]. The combined signal and pump energy propagate along the EDF, where signal amplification occurs, and then the amplified signal exits the EDF and passes through a second isolator. The purpose of the two isolators, which allow light to pass only in a single direction, is to ensure that lasing cannot take place within the EDF. Furthermore, the output isolator also acts as a filter for 980nm light propagating in the forward direction, thus stopping the 980nm light from exiting the amplifier output port[4].

A. Type of Pumping and Its Simulation

Basically there are three types of pumping methods used in erbium-doped fiber amplifier

B. Forward Pumping

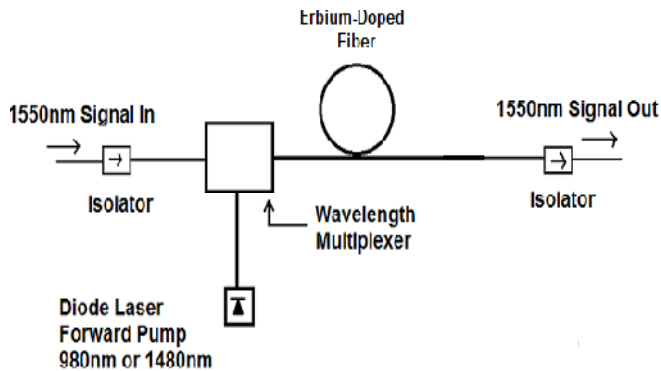


Fig. 1.2 Structure of forward pumping

In forward pumping the input signal and the pump signal propagate in the same direction inside the fiber. The input signal and pump are combined using a pump combiner or

wavelength division multiplexer. Inside the fiber the pump energy is transferred to the input signal and the signal is amplified at the output of the amplifier. Isolators are used in the scheme to make sure that the signal will travel only in one direction and no feedback of signal will occur.

C. Backward Pumping

In Backward pumping the input signal and the pump signal propagate in the opposite direction to each other inside the fiber. For amplification the direction of input and pump signal is not essential. They can travel in any direction.

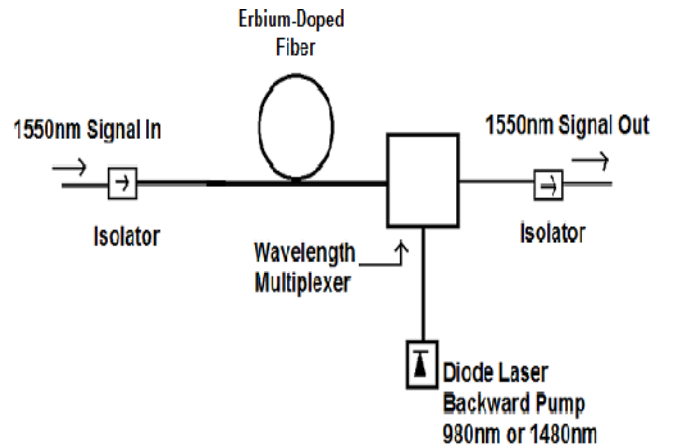


Fig. 1.3. Structure of backward pumping

D. Bi-Directional Pumping

In Bi-directional pumping the input signal travels in one direction. But there are two pump signals that travel inside the fiber. One pump signal travels in the same direction as the input signal and the other pump signal travels in the opposite direction to that of the input signal.

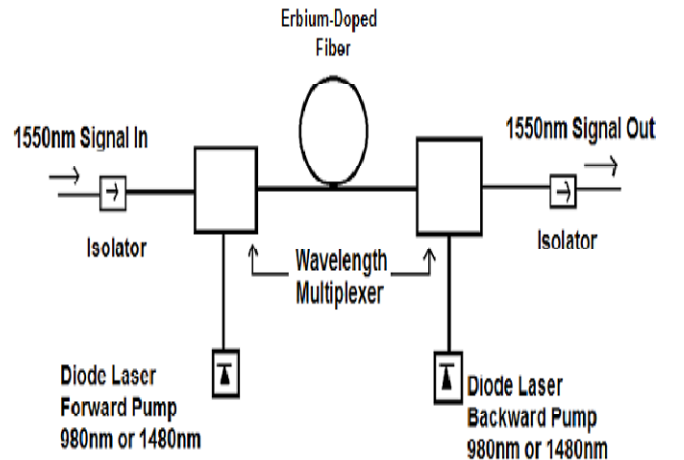


Fig. 1.4 Structure of bi-directional pumping

E. Simulation of Pumping

Simulation of erbium based amplifier done for different length. length varies from 5 to 50 m the input signal is -30 dBm for 1550 nm meter wavelength. the input pump power is 200 mW [5].

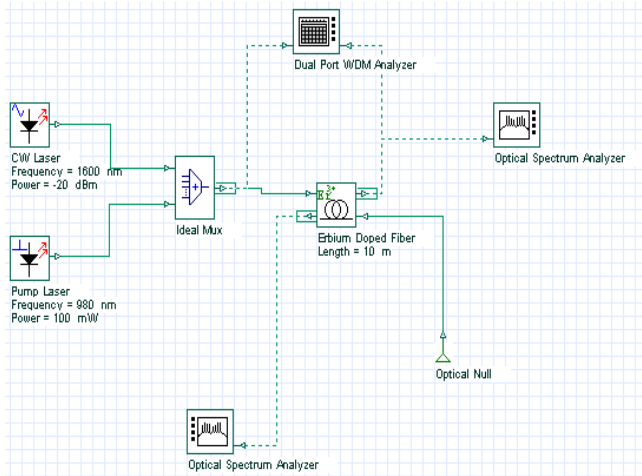


Fig. 1.5 Simulation of forward pumping @960 mW pump power

4. RESULTS AND DISCUSSIONS

We will study the different value of gain for different length of fiber for different pumping scheme. It was observed that gain for 980 nm pump is almost same for forward and backward pumping and for bi directional pumping gain is quite high. The similar case happened for 1480 nm pumping wavelength. Noise figure for co-pump is the lowest at both 980 nm and 1480 nm pumping [6]. For counter pumping Noise Figure is the highest among all the pumping method.

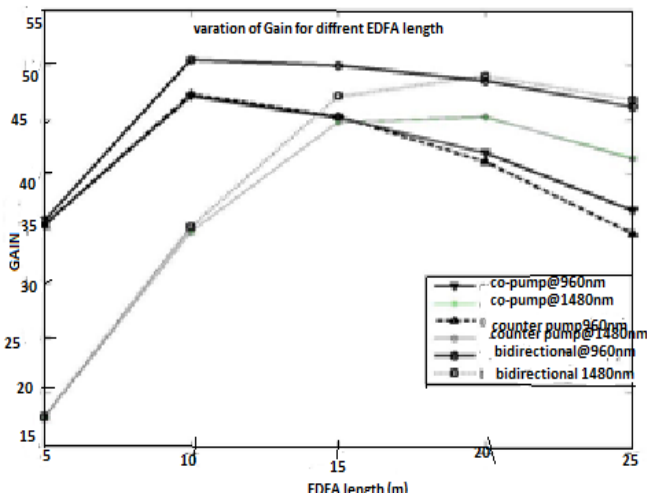


Fig. 1.6 Variation of gain with length of fiber for different pumping methods

The above figure shows the variation of gain for different pumping methods for different length of erbium doped fiber; we take consideration of fiber length 5 to 25 m and studied the variation of gain for different pumping methods [9]. We can observe that for constant gain we can choose the co-pump because after the length of fiber increased from 10 to 15 almost constant gain can be obtained with better value. For the low short length of fiber bidirectional pumping only provided lower value of gain hence these pumping is used with high value of pumping power and larger length of fiber.

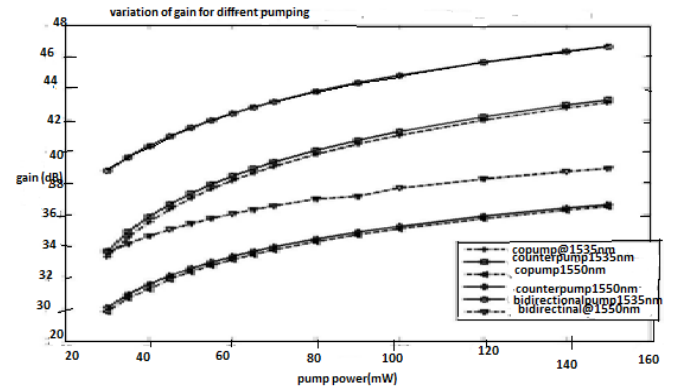


Fig. 1.6 Variation of Gain w.r.t pumping power for different pumping methods

In fig. 1.6 for 1535 nm signal is higher for counter pump than co-pump than 1550 nm signal, because the 1535 nm signal has higher cross section than 1550 nm signal. We see in the above diagram gain and noise figure for different value of erbium fiber and for different wave length [7].

We can see in the figure the better noise figure (NF) can be found by using co-pumping method, but in case of better amplification this can be found by using bi-directional pumping [8].

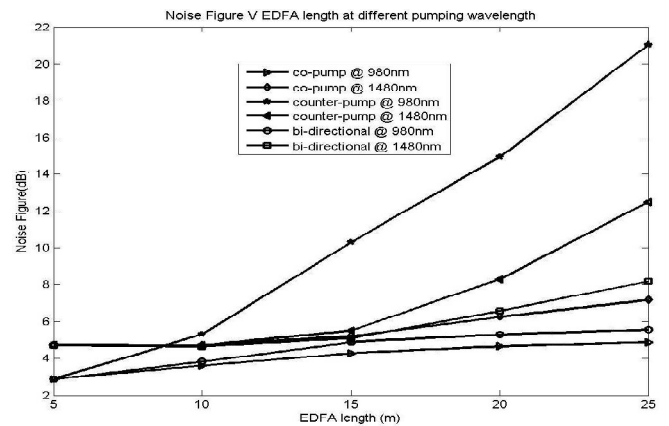


Fig. 1.7. Variation of Noise Figure (NF) w.r.t length of fiber for different pumping power

Bi-directional directional pumping is also useful in case of low noise figure and better amplification. But when we concentrate only on to get better gain value this can be achieved by using counter pumping methods [10].

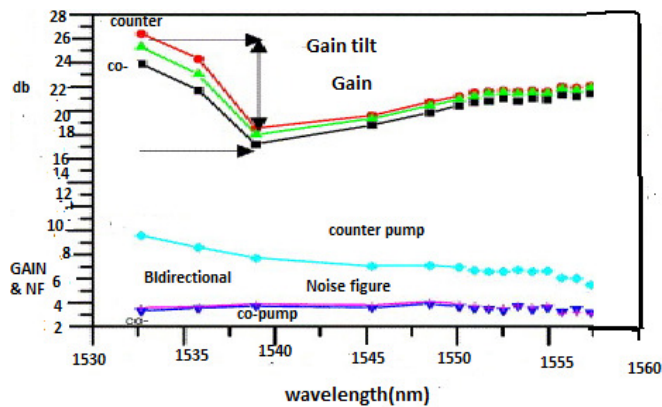


Fig. 1.8. Comparative analysis of Noise Figure (NF) and Gain for different wavelength and different pumping methods

5. CONCLUSIONS

The Objective of this paper was the comparative analysis of different methods of pumping for erbium doped fiber amplifier (EDFA) for different parameter such as gain noise figure and pumping power is achieved by using optisystem 11.0 .It is observed by the experimental determination that bi-directional pumping is the best way of achieving high power amplification with moderate noise level in EDF fiber. It is found that for different situation different pumping technique is applicable .By changing the parameter in

pumping technique the gain and noise figure also changed. It is found from graphical analysis when better noise figure demanded we use co-pumping, for better gain and noise figure and better amplification we use bi-directional pumping.

REFERENCES

- [1] R.Deepa, R.Vijaya "influence of bidirectional pumping in high -power EDFA on single -channel, multichannel, and pulsed signal amplification" ELESIVER.
- [2] [Rajesh Kaler, R.S Kaler" Gain and noise figure performance of erbium doped fiber amplifiers (EDFAS) and compact EDFAS"ELSVIER OPTIK2011.
- [3] M.A Othman, et.al"Erbium doped fiber amplifier (EDFA) for c-band optical communication "JET-IJNES VOL;12 NO;04 2012.
- [4] A.Cem Cokrak,Ahmet Altuncun" Gain and noise figure performance of erbium doped fiber amplifier(EDFA)JEEE 2004
- [5] Optisystem component library2011.
- [6] Bannaz O.Rashid, perykhan.M.Jaff"Gain and Noise figure performance of erbium -doped fiber amplifier at 10GPS".2010.
- [7] M. Pal, M.C. Paul, A. Dhar, A. Pal, R. Sen, K. Dasgupta and S.K. Bhadra," Investigation of the optical gain and noise figure for multichannel amplification in EDFA under optimized pump condition", pp.407-412, Elsevier, 2007.
- [8] White paper "introduction of EDFA technology" 2009.
- [9] Emmaunuel Desurvire" Erbium Doped Fiber Amplifiers" principles and Application, Republished in New York by Colombia University 1994.
- [10] K. Bertilsson, P.A. Anderkson, J. Lightwave technology.12 (7) (1994)1198-1206.

Mode Delay of Optical Waveguide Using Equivalent TL Circuit

Vinita Mathur¹, Parul Tyagi², Surbhi Jain³

^{1,2,3}ECE, JECRC Jaipur

¹vinitamathur12@gmail.com, ²tyagi.parul82@gmail.com, ³surbhi.first@gmail.com

Abstract: In this paper an explicit program has been presented in order to directly calculate the mode delay of cylindrical waveguide of erratic refractive-index profile. Various fiber substantial's have been used and depending on their permittivity there delays have been computed. MATLAB is used as a tool for the estimation of the delays, using commensurate transmission-line (T-L) technique. Using Maxwell's equation, an analogous T-L circuit for a cylindrical dielectric waveguide has been derived [1]. A recursive program has been written using the T-L model which permits calculation of mode delay directly. There is a large number of research papers on application of TL method of waveguide analysis. [2, 3]. The recommended program without the need for curve fitting and subsequent successive numerical differentiation calculates delay directly from the propagation constant. It is literal, rapidly imminent, and it results in savings for both storage memory and computing time.

Keywords: Interconnects, modes, optical fiber, line impedances, dispersion, propagation constants, delay.

1. INTRODUCTION

In today's era interconnect has appeared essential impediment in integrated circuit design. With the scaling of CMOS technology copper interconnects would not be able to satiate the design requisite of delay, power, bandwidth and noise. From the past two decades electrical interconnects has been replaced by on-chip optical interconnects.

In the field of telecommunication optical devices are primarily used and are also regularly applied as board level interconnects. In 1984 Goodman first introduced the theory of on chip optical interconnects [4]. Since electrical/optical and optical/electrical modification is required, optical interconnects is specifically comely for global interconnects, such as data buses and clock distribution networks. Anew, several analyses have been made on chip electrical and optical interconnects. This analysis is principally assert as optical interconnect is a active developing technology whereas electrical interconnect is relatively developed. Despite the cognizance of on chip optical interconnect lean on the development of enhanced CMOS compatible optical devices [5].

2. ANALYSIS

Computation of waveguide mode propagation constants as a function of wavelength for optical fibers is a well authorized issue. In our method we have calculated total mode delay from the mode propagation constants. The terminology of dispersion includes the use of first (delay) and second (dispersion) derivatives of mode propagation constant with respect to wavelength, therefore theoretical calculation of dispersion needs the determination of such derivatives in the first sample.

We have shown that equivalent transmission line(TL) circuit technique are most powerful and can be easily applied to optical fibers in order to determine exactly the mode propagation constants.

In this paper a procedure based on the TL circuit technique for evaluating the mode delay of optical fibers of known but any random refractive index profiles are presented.

Step1- We derive the equations for the derivatives of the propagation constant with respect to wavelength.

Step2- With the use of recursive formula, we have shown that for a given wavelength, the first derivative can be expressed in terms of equivalent circuit impedances at the wavelength of interest.

Step3- Finally by calculating the derived equivalent circuit formulas, we are able to work out the impedances at the specified wavelength using the T-L technique. The total delay can be accurately calculated.

3. TRANSMISSION LINE THEORY

Our analysis divides a cylindrical symmetric optical fiber into a large number of concentric homogeneous cylindrical layers of thickness δr permittivity ϵ , permeability μ , and conductivity σ in fig 1.

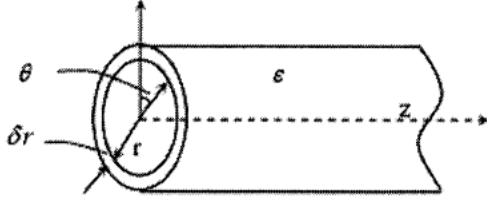


Fig. 1. Homogeneous optical fiber thin cylindrical layer

Using Maxwell's equations for the E and H fields we derive the following equations for any such layer

$$\beta r E_\theta - I E_z = w \mu_r H_r$$

$$I H_z - \beta_r H_\theta = (w \epsilon - j \sigma) r E_r$$

$$\left(\frac{\partial[(\omega \epsilon - j) r E_r]}{\partial r} = -(\sigma + j \omega \epsilon)(I E_\theta + \beta_r E_z) \right)$$

$$\frac{\partial(I H_\theta + \beta_r H_z)}{\partial r} = -\frac{\gamma^2}{j \omega \mu} w \mu_r H_r + \beta H_r - \frac{1}{r} H_\theta$$

$$\frac{\partial(I E_\theta + \beta_r E_z)}{\partial r} = -\frac{\gamma^2}{\sigma + j \omega \epsilon} (\omega \epsilon - j \sigma) r E_r + \beta E_r - \frac{1}{r} E_\theta \quad \text{-----2}$$

where β = propagation constant

n_{eff} = effective refractive index

l = interger azimuthal mode number

w = mode frequency

ϵ_r = material's relative permittivity

μ_r = relative permeability

Where

$$\gamma^2 = \beta^2 + \frac{1}{r^2} - w^2 \mu \epsilon + j w \mu \epsilon + j \omega \mu \epsilon$$

In case of light, the refractive index (n) of the layer at a distance r from the axis, it equals

$$n = \sqrt{\epsilon_r \mu_r}$$

$$V_s = \frac{V_m}{\sqrt{n}} + V_E \sqrt{n}$$

$$V_d = \frac{V_m}{\sqrt{n}} - V_E \sqrt{n}$$

$$I_s = I_M \sqrt{n} + \frac{I_e}{\sqrt{n}}$$

$$I_d = I_M \sqrt{n} - \frac{I_e}{\sqrt{n}}$$

Where

$$V_M = \frac{I H_\theta + \beta_r H_z}{j F} Z_\theta \quad (\text{magnetic voltage})$$

$$V_E = \frac{I H_\theta + \beta_r H_z}{j F} Z_\theta \quad (\text{electric voltage})$$

$$I_E = \omega \epsilon_0 n^2 r E_r \quad (\text{electric current})$$

After some algebra [6] (1) and (2) can be transformed into

$$\left(\frac{\partial V_s}{\partial r} = \frac{\gamma^2}{j \omega \epsilon_0 n F} I_s \right)$$

$$\frac{\partial I_s}{\partial r} = -j \omega \epsilon_0 n F V_s \quad \text{-----3}$$

$$\frac{\partial I_d}{\partial r} = -j \omega \epsilon_0 n F V_d \quad \text{-----4}$$

where

$$\gamma^2 = \beta^2 + \left(\frac{l}{r} \right)^2 - (nk)^2 + \frac{2nk\beta l}{(\beta r)^2 + (l)^2}$$

$$F = \frac{(\beta r)^2 + (l)^2}{r}$$

Equations 3 and 4 represent two independent transmission lines with voltages V_s and V_d and currents I_s and I_d . The corresponding characteristic impedances are

$$Z_s = \frac{\gamma_s}{j \omega \epsilon_0 n F} \quad Z_d = \frac{\gamma_d}{j \omega \epsilon_0 n F}$$

Equations (3) and (4) are recognized as T-L equations the solution of which can be represented by the following equivalent electric circuit, fig 2

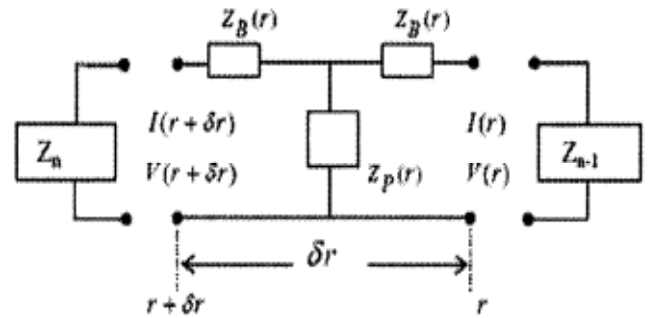


Fig. 2. Equivalent circuit of a dielectric waveguide layer

The transmission line impedances are given by, were
 $Z_B = \sinh(\gamma_d \delta r) \tanh(\gamma_d * 0.5 * \delta r) Z_p$

$$Z_p = \frac{\gamma_d Z_0}{j n r k o \left(\beta^2 + \left(\frac{l}{r} \right)^2 \right) \sinh(\gamma_d \delta r)}$$

An optical fiber can be represented as a cascade of TL circuits connected in tandem.

The equivalent TL circuit impedances (7) or (8) are functions of wavelength and the propagation constant, so the first derivative of the propagation constant can be extended as follows:

$$\frac{\partial \beta}{\partial \lambda} = \frac{\left(\frac{\partial Z_n}{\partial \lambda}\right) \beta = \beta_o}{\left(\frac{\partial Z_n}{\partial \beta}\right) \lambda = \lambda_o}$$

$$Z_n = Z_B + \frac{1}{Z_B + Z_{n+1}} + \frac{1}{Z_p} - 1$$

($n=1,2,\dots,N$), which is the n th characteristic impedance of cylindrical layers. N is the total number of cylindrical layers.

4. SOLUTION PROCEDURE

The material refractive index dependence [7] on wavelength included in calculations is given by

$$n_1(\lambda) = C_0 + C_1\lambda^2 + C_2\lambda^4 + \frac{C_3}{\lambda^2 - 0.035} + \frac{C_4}{(\lambda^2 - 0.035)^2} + \frac{C_5}{(\lambda^2 - 0.035)^3}$$

Where $C_0 = 1.4508554$, $C_1 = -0.0031268$, $C_2 = -0.0000381$, $C_3 = 0.0030270$, $C_4 = -0.0000779$, $C_5 = 0.0000018$.

Delay equation is given by

$$\tau = \frac{L\lambda^2}{2 \times 3.14 \times c} \times \frac{\partial \beta}{\partial \lambda}$$

Where τ is the delay, L is the optical fiber length, and c is the velocity of light in free space.

5. COMPARISON

Here we have considered different fiber materials with different permittivity. We have taken wavelength as 1550×10^{-9} , azimuthal mode number as 1, core radius as 1×10^{-6} , length of the fiber as 1×10^{-2} and calculated the delay on MATLAB program.

Table 1

Material	Wavelength	Permittivity	Delay
SiO ₂	1550×10^{-9}	3.9	3.6286×10^{-14}
AlAs	1550×10^{-9}	10.1	5.8028×10^{-14}
Si	1550×10^{-9}	11.9	6.2898×10^{-14}
Ge	1550×10^{-9}	16	7.2725×10^{-14}
BaSrTiO ₃	1550×10^{-9}	200	2.4436×10^{-13}

6. CONCLUSION

In this paper, a precise program for evaluating the mode delay of cylindrical waveguide has been developed. This method uses T-L representation of cylindrical waveguides and relies on the modeling of a thin uniform concentric cylindrical layer of an optical fiber to a T-L circuit. The method requires knowledge of only the mode propagation constant and the refractive index profile. It is direct and exact, and avoids the use of numerical differentiation twice. It may be especially useful for designing and predicting complex refractive index profile optical fibers where the earlier reported approximate methods are quite slow.

REFERENCES

- [1] A.C. Boucouvalas and Xia Qian., "Mode Dispersion and delay characteristics of optical waveguides using equivalent TL circuits", Quantum Electronics, IEEE Journal, 41(7) July 2005, 951-957.
- [2] S.K. Raghuwanshi, V. Kumar and R.R. Pandey, "Derivation of Eigen value equation by using equivalent TL method for the case of Symmetric/ Asymmetric Planar Slab Waveguide Structure", Vol. 15 No.1 (2011), Journal of International Academy of Physical Sciences, pp. 113-122.
- [3] Xin Qian and A.C. Boucouvalas, "Analysis of leaky modes and Bragg fibers using transmission line equivalent T-circuits, IEEE Photonics Tech. Lett., 17(5) (2005) 1031-1033.
- [4] J.W. Goodman *et al.*, "Optical Interconnects for VLSI Systems," *Proceedings of the IEEE*, Vol.72, No. 7, PP. 850-866, July 1984.
- [5] Guoqing Chen and David Albonesi, "Predictions of CMOS Compatible On-Chip Optical Interconnect," April 2005
- [6] X. Qian and A.C. Boucouvalas, "Optical fiber refractive index profile synthesis from near field," in *Proc. IEEE Globecom Conf.*, San Francisco, CA, 2003, pp. 2669-2673.
- [7] S. P. Survaiya and R. K. Shevgaonkar, "Dispersion characteristics of an optical fiber having linear chirp refractive index profile," *J. Lightw. Technol.*, vol. 17, no. 10, pp. 1797-1805, Oct. 1999.

Metamaterials: A New Boon in Design of Frequency Selective Surfaces

Reepika Sharma¹, Ankush Kapoor²

¹B.Tech 2nd Year ECE Department, Jawaharlal Nehru Government Engineering College
Sundernagar Distt. Mandi H.P, Pin-175018, reepika07@gmail.com

²Assistant Professor ECE Department, Jawaharlal Nehru Government Engineering College
Sundernagar Distt. Mandi H.P, Pin-175018, ankush8818@yahoo.com

Abstract: This paper presents the importance and necessity of smart antenna for mitigation of multipath fading in typical wireless scenarios and to improve the diversity gain, capacity and overall performance of the system. We have tried to explain the importance of Smart Antennas in the present era.

Keywords: Metamaterial, Frequency Selective Surface (FSS), Spatial Filters

1. INTRODUCTION

Electronic circuit technologies have been developed with the support of the research and development of many materials including dielectrics and magnetic substances. Meanwhile, there seems to be a growing number of attempts to realize electromagnetic characteristics by means of artificial structures called “metamaterials” instead of the characteristics that have conventionally been gained from the physical properties of materials [1]. Metamaterials are materials which exhibit properties that may not be found in nature. They are ensemble of multiple individual elements of conventional microscopic materials such as metals or plastics, but the materials are usually arranged in periodic fashion [2-3]. An array of periodic metallic patches on a substrate, or a conducting sheet periodically perforated with apertures, constitutes a frequency selective surface (FSS) to electromagnetic waves. Such structures have been well known in antenna theory for over half a century. At microwave wavelengths, such structures were easy to manufacture and employ in antenna design. However, at near-infrared wavelengths, the sizes of elements in the FSS are on the order of a micrometer making such structures much more difficult to manufacture.

2. DESCRIPTION OF FSS

A conducting sheet periodically perforated with apertures, or an array of periodic metallic patches, constitutes a frequency selective surface (FSS) to electromagnetic waves. In the literature two generic geometries are typically discussed. The first geometry, commonly referred to as an inductive FSS, performs similarly to a high-pass filter. The second case, or capacitive FSS, is similar to a low-pass filter. If the

periodic elements within an FSS possess resonance characteristics, the inductive FSS will exhibit total transmission at wavelengths near the resonant wavelength, while the capacitive FSS will exhibit total reflection [4,5,7]. Capacitive and inductive FSSs derive their name from circuit theory. Figure 1 shows a typical capacitive and inductive FSS constructed out of periodic rectangular patches and apertures respectively. Also included in the figure are their respective equivalent circuit models, along with the corresponding transmission profiles.

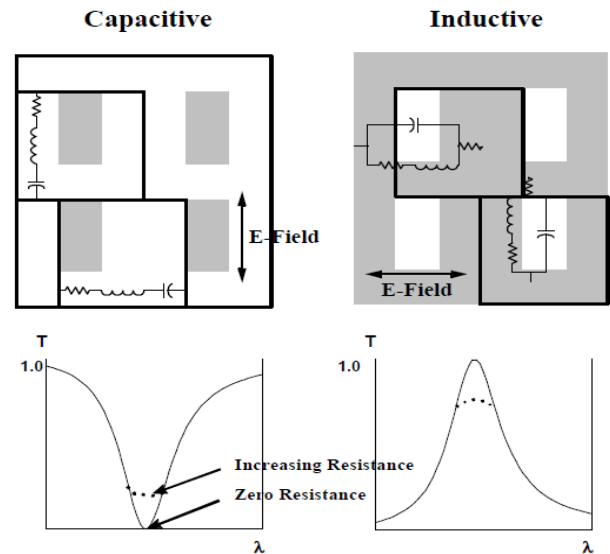


Fig. 1. Capacitive and inductive FSSs with corresponding equivalent circuits and their transmission profiles. Note that only 4 of the periodic elements per filter are shown in the drawing.

Typical FSSs for the near infrared region will have hundreds of thousands of periodic elements[6].

The rectangular metallic patches in the capacitive FSS act similar to a capacitive circuit. Similarly, the rectangular apertures in the inductive FSS behave like an inductive circuit. The FSS periodic surface may be planar, or it may have a profile in the third dimension. Such three dimensional periodic gratings are commonly referred to as Surface Relief

Gratings [8]. This dissertation deals exclusively with planar FSSs. Typical lithographic techniques such as the one described in Figure 2 will produce a planar metallic filter surface deposited on a substrate. A capacitive type filter will consist of metallic patches deposited on a planar substrate. For the inductive type filter, a metallic sheet, usually deposited on a substrate, is perforated with apertures.

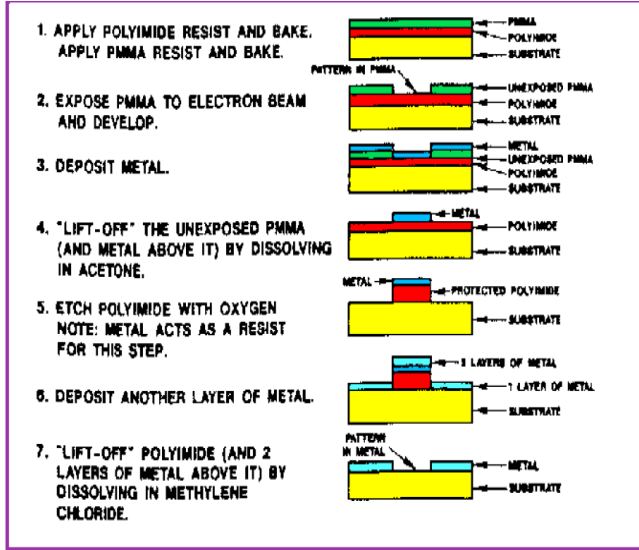


Fig. 2. Typical lithographic process for creating a planar FSS on a substrate [4].

The periodic elements in a FSS are most commonly arranged in a rectangular array as shown in Figure 3. However, the more general geometric arrangement is a triangular array, also shown in figure 3. Note that the periodicity in the triangular array exists along the x -axis, and the skewed $-x$ -axis. If the skew angle $\alpha=90^\circ$, then the triangular array becomes a rectangular array [9,10,11]. The geometric labeling shown in Figure 3 will be used throughout this dissertation.

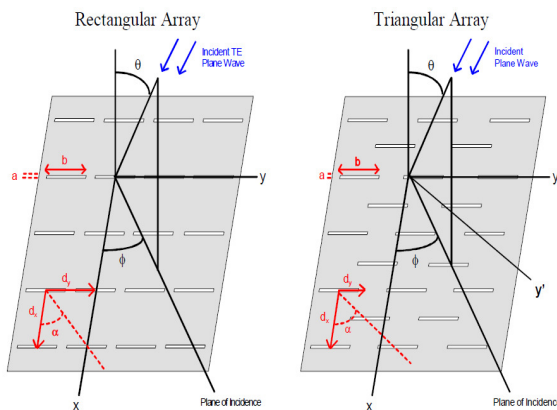


Fig. 3. A rectangular array and the more general triangular array of apertures (or patches) that form a FSS. Note that when the skew angle $\alpha=90^\circ$, the triangular array becomes a rectangular array [11].

It should be pointed out that the incident field is considered to be a plane wave. The angle of incidence and polarization is not restricted in the FSS model used in this research. Polarization states are typically divided into the orthogonal TE (electric field perpendicular to plane of incidence) and TM (magnetic field perpendicular to plane of incidence) states.

3. APPLICATIONS OF FSS

In the near infrared wavelength region, frequency selective surfaces are primarily used in filtering applications. Some applications include bandpass filters, polarizers, and beam splitters. As previously stated, if the periodic elements within a FSS possess resonance characteristics, the inductive FSS will exhibit total transmission at wavelengths near the resonant wavelength, while the capacitive FSS will exhibit total reflection [14, 15]. This feature allows an FSS with the proper elements to perform like a narrow bandpass filter.

Similarly, a complimentary design can be created to selectively absorb narrow bandpass regions. Specific applications of these types of FSS filters include narrowband astronomy filters, and filters for spacecraft instrumentation. If the periodic elements within a FSS possess a resonance characteristic that is polarization dependent, such a feature may be exploited to produce a polarizer. If this polarization dependence is utilized so that one polarization is totally reflected, while the orthogonal polarization is totally transmitted, the FSS may then be employed as a beam splitter. Using FSSs allows construction of polarizers and beam splitter for wavelengths where traditional materials make construction of such devices impractical.

4. METAMATERIAL

Periodically arranged at intervals shorter than the specified wavelength of an electromagnetic wave, small pieces of metal and the like can constitute an artificial medium that has characteristics not found in nature (Fig. 1). Such a medium is called metamaterial. Metamaterial can also be made of dielectrics, magnetic substances, semiconductors, and the like, and even electric circuits instead of metal pieces. The word “meta” derives from the Greek word that means “beyond.” While conventional materials provide their intended physical properties in terms of design on the atomic or molecular level, metamaterials realize their specified physical properties through the design of an artificial structure that can be regarded as a quasi-uniform medium in a macroscopic view.

4.1 Left Handed Metamaterial

The electromagnetic characteristics of electronic material are primarily determined by the basic parameters, i.e. the permittivity, the permeability, and the conductivity. For

example, passive components such as capacitors and inductors are typically made of materials where both the permittivity and the permeability are positive. These materials are referred to as right-handed materials since the vectors of the electric field, the magnetic field, and the wave number of the electromagnetic wave in the materials correspond in direction to the thumb and two fingers of the right hand (the first quadrant in Figure 4).

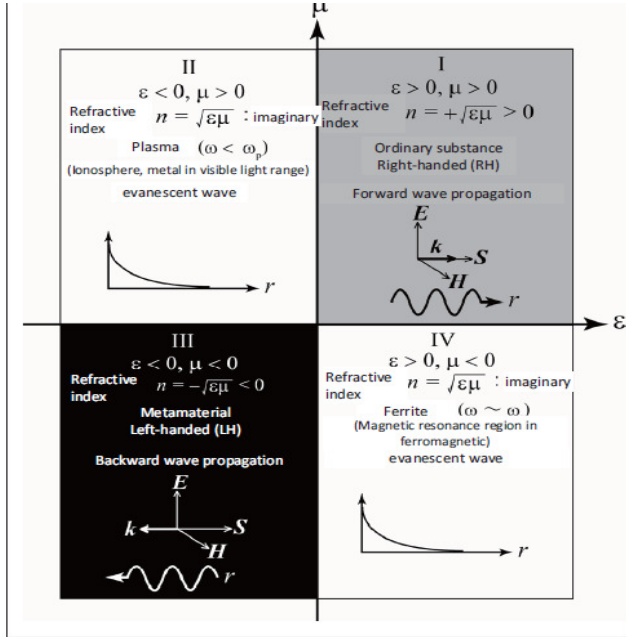


Fig. 4. Classification of Materials by Permittivity ϵ and Permeability μ [3].

In contrast to such ordinary electronic materials, ones with this simultaneously negative-permittivity and permeability, if any, are referred to as left-handed materials since the vectors correspond in direction to the thumb and two fingers of the left hand (the third quadrant in Figure 4). There exist no left-handed materials in nature, however. Left-handed materials produce peculiar phenomena, among which a “negative refractive index” and the generation of a “backward wave” are the properties of particular significance.

4.1.1 Categories of Left Handed Metamaterial

The paper [3] of a metamaterial, i.e., a left-handed material consisting of an artificial structure, in 2000 sparked research into “left-handed metamaterials” for practical use. A left-handed metamaterial is an artificial structure in which small pieces of metal or the like are periodically arranged at an interval shorter than the wavelength of the intended electromagnetic wave. Each individual portion of the periodical structure is called “a unit cell”. The left-handed metamaterial is fabricated by optimizing the shape and

arrangement of the unit cells so that the artificial structure has the aimed characteristics. Among the characteristics produced from these artificially-structured metamaterials, left-handed metamaterials are regarded as a technique to make positive use of “dispersion characteristics” that change with frequency.

In other words, left handed metamaterials inevitably have frequency dependencies and show left-handed characteristics in a certain frequency band. It follows that left handed metamaterials may also show right-handed characteristics or rejection characteristics as well in other frequency bands. In the field of information and communications, many left-handed metamaterials are used in many applications as a combination of left handed and right-handed elements, rather than as sole left-handed elements, and the former applications are the more dominant in practice. Such metamaterials are sometimes referred to as CRLH (Composite Right/Left-Handed) metamaterials.

4.1.2 Examples of Applications of Left Handed Metamaterial

1. Wide-Range Beam Scan of Radar Antenna:

Radar is a system that performs a wide range scanning detection across a certain angle by rotating its antenna having a directional narrow beam radiation pattern. The phase of the electromagnetic wave inside the antenna is so controlled as to implement the antenna's beam scan without using a rotation mechanism. A backward wave from a left handed metamaterial can be used effectively to widen the range of beam scanning. Sweeping the frequency of the radar electromagnetic wave from the right handed frequency band to the left-handed frequency band, both the ordinary forward wave (electromagnetic wave in a right-handed medium) and the backward wave can be used.

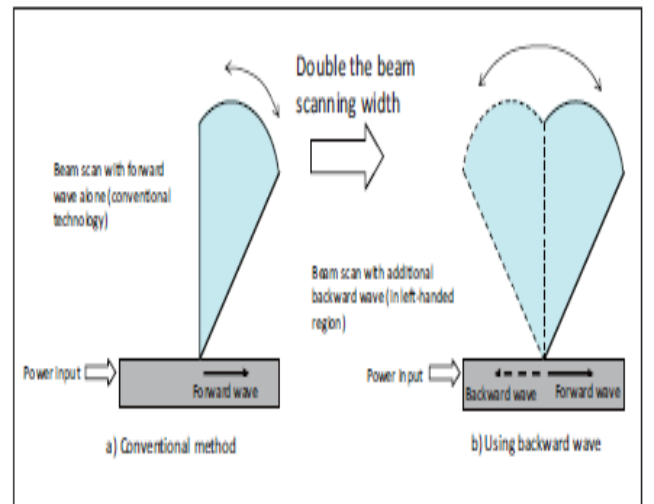


Fig. 5. Practical Example of Beam Scan of Radar Antenna [2]

2. Adjustment of system characteristics:

The characteristics of a left-handed transmission line can be utilized to provide a device that has a special effect applicable to electromagnetic appliances. Figure 6 shows devices in which two transmission lines are juxtaposed to each other. Such devices, called “couplers,” have the function of transferring electromagnetic waves that propagate in each of the transmission lines to the another for power distribution or transmission, and are in use for various power handling purposes [3,5]. A backward coupler, which includes a left-handed transmission line, enables the adjustment of system characteristics for improved system performance and space saving. A coupler with ordinary right-handed transmission lines (Figure 6) splits and transmits the electromagnetic wave in the same direction, i.e., from port 1 to ports 2 and 3.

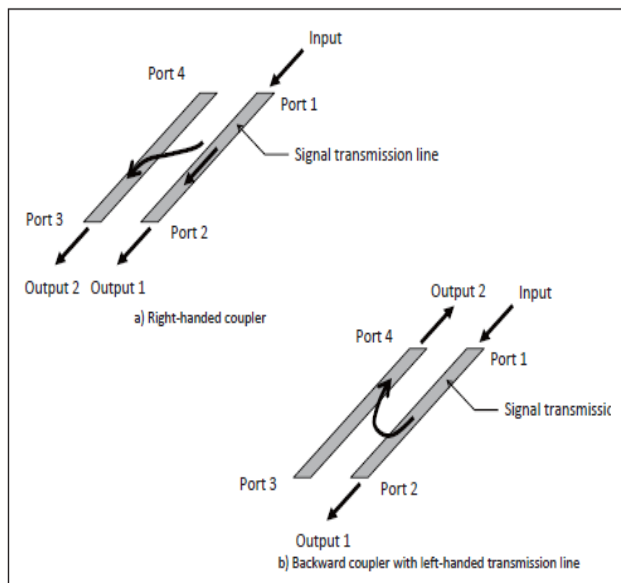


Fig. 6. Backward Coupler [1]

3. Reduction of Electromagnetic Interference

A transition region or a stop band between left- and right-handed frequency bands may be used to isolate or attenuate undesired electromagnetic waves. This effect is expected to be applied to circuit integration for cellular phones that incorporate a plurality of radio systems and digital circuits. For example, the current cellular phone system employs the 800- MHz, 1500-MHz, and 2-GHz bands for phone use, the 1.57-GHz band for GPS, the 470- to 710-MHz bands for 1seg TV, and the 13.56-MHz band for electronic payment and other applications. A plurality of antennas for covering the respective specific frequency bands are mounted in the limited available space. Communication performances are degraded by electromagnetic interferences between the antennas inside installed in a small space.

4. Metamaterial – A New Design Approach to FSS

Traditional frequency-selective surface (FSS) structures, with resonant unit cells, have been investigated over the years for a variety of applications. These include bandpass and bandstop spatial filters, absorbers, and artificial electromagnetic bandgap materials. A typical FSS is a 2-D planar structure consisting of one or more metallic patterns, each backed by a dielectric substrate. These structures are usually arranged in a periodic fashion; therefore, their frequency response is entirely determined by the geometry of the structure in one period called a unit cell. As a result of research on the applications mentioned above, the behavior of FSSs is well understood [7]. The focus of the past studies, however, has been mostly on the bandstop characteristics produced by these surfaces, and structures with bandpass characteristics have been rarely studied. Recently, there has been an interest in design of FSS with unit cell dimensions much smaller than a wavelength. In traditional designs, the frequency-selective properties result from mutual interactions of the unit cells. Therefore, to observe a desired frequency selective behavior, a large number of unit cells must be present. Consequently, the overall size of the surface is electrically large. On the other hand, for some applications where a low sensitivity with respect to the incidence angle of the exciting wave is required or in cases where a uniform phase front is difficult to establish, the screen size needs to be small. To address this problem, a new class of FSSs called miniaturized-element frequency-selective surfaces was developed [8]. The new class takes an approach which is different from those of the past designs. In this approach, instead of using a resonant structure as the building block of the frequency selective surface, special unit cells of small dimensions are used. These unit cells act as lumped inductive and capacitive elements and are properly arranged so they couple to the magnetic and electric fields of an incident wave, respectively.

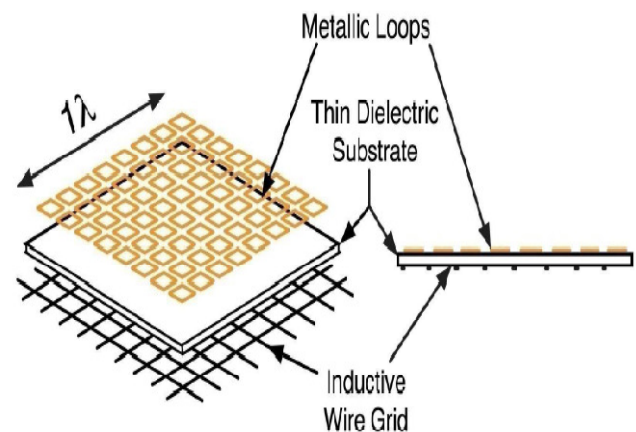


Fig. 7. Metamaterial based FSS

5. CONCLUSION

Frequency Selective Surfaces are still an ongoing research field and this article will prove to be beneficial for the researchers to develop FSS with more accuracy as we have introduced a new concept based on metamaterials.

A new theory in design of microwave, spatial filters for compact communications applications is the main advance of this review. So, by studying the effect of the substrate thickness, the design of the Frequency Selective Surface may be adjusted to give the desired transmission and reflection characteristics.

REFERENCES

- [1] T. K. Wu, "Frequency Selective Surface and Grid Array", New York: Wiley, 1995.
- [2] B. A. Munk, "Frequency Selective Surfaces: Theory and Design", New York: Wiley, 2000, pp. 2–23.
- [3] R. Mittra, C. H. Tsao, and W. L. Ko, "Frequency selective surfaces with applications in microwaves and optics," in Proc. IEEE Microwave Symp, vol. 80, pp. 447 – 449, 1980.
- [4] Abbaspour-Tamijani, K. S., and G. M. Rebeiz, "Antenna Filter-Antenna Arrays as a Class of Band-Pass Frequency Selective Surfaces", IEEE Trans. Microw. Theory Tech., vol 52 , pp . 1781 -1789, 2004.
- [5] K.Sarabandi and N. Behdad, "A Frequency Selective Surface with Miniaturized Elements," IEEE Trans. Antennas Propag., vol 55, pp . 1239-1245, 2007.
- [6] N. Marcuwitz, "Waveguide Handbook" (1st edition), NewYork: McGraw-Hill, 1951.
- [7] R.J. Langley, E.A. Parker, "Double square frequency selective surfaces and their equivalent circuit", Electronics Letters, vol. 19, no. 17, pp. 675 – 677, 1983.
- [8] E.F. Kent, B. Doken and M. Kartal "A New Equivalent Circuit Based FSS Design Method by Using Genetic Algorithm", International Conference on Engineering Optimization, Lisbon, Portugal, Sept. 2010.
- [9] L. Zappelli, "Analysis of modified dielectric frequency selective sur-faces under 3-D plane wave excitation using a multimode equivalent network approach," IEEE Trans. Antennas Propag., Vol. 57, No. 4, pp. 1105–1114, April 2009..
- [10] F. Bayatpur, K. Sarabandi, "Single-Layer, High-Order, Miniaturized-Element Frequency Selective Surfaces," IEEE Transactions on Microwave Theory and Techniques, Nov. 2008.
- [11] Y.J Lee, J. Yeo, R. Mittra, and W. S. Park, "Application of electromagnetic bandgap (EBG) superstrates with controllable defects for a class of patch antennas as spatial angular filters", IEEE Trans. Antennas and Propag. , vol. AP-53, no. 1, pp. 224–234, Jan. 2005.
- [12] S. Biber, M. Bozzi, O. Gunther, L. Perregrini, and L. P. Schmidt, "De-sign and testing of frequency-selective surfaces on silicon substrates for submillimeter-wave applications," IEEE Trans. Antennas Propag., Vol. 54, No. 9, pp. 2638–2645, 2006.
- [13] J.P. Gianvittorio, J. Zendejas, Y. Rahmat-Samii and J. Judy, "Reconfigurable MEMS-enabled frequency selective surfaces," IEEE Electron.Lett., Vol. 38, No. 25, pp. 1627–1628, Dec. 2002.
- [14] M. Ohira, H. Deguchi, M. Tsuji, and H. Shigesawa, "Multiband single-layer frequency selective surface designed by combination of genetic algorithm and geometry-refinement technique," IEEE Trans. Antennas Propag., vol. 52, no. 11, pp. 2925–2931, Nov. 2004.
- [15] R.A.Hill and B.A. Munk, "The effect of perturbing a frequency selective surface and its relation to the design of a dual-band surface," IEEE Trans. Antennas Propagat., vol. AP-44, no.3, pp.368-374, Mar. 1996.
- [16] A.E. Yilmaz and M. Kuzuoglu, "Design of the square loop frequency selective surfaces with particle swarm optimization via the equivalent circuit model, vol. 18, no. 2, pp. 95-102, 2009.

Performance Analysis of OCDMA System using Various Amplifiers

Charu Singh¹, Brahmraj Singh², Brijesh Jaiswal³

^{1,2,3}School of ICT, Gautam Buddha University, Greater Noida, INDIA

¹charusingh1301@gmail.com, brahmrajs35@gmail.com, brijeshjaiswal13011@gmail.com

Abstract: Optical fiber communication deals with higher data rate through different types of optical fibers. OCDMA technique uses optical fiber as a medium for transfer of data from source to destination. So amplifiers used for signal amplification plays an important role in data transmission. This paper basically deals with use of different types of amplifier used in OCDMA system. Optical Amplifier, EDFA with FRA is being used in OCDMA system. The various parameters like BER, Q-factor and Eye diagram are being analyzed for different amplifiers.

Keywords: OCDMA, EDFA, FRA, BER, Q Factor.

1. INTRODUCTION

Optical code division multiple access (OCDMA) is a technique which is used to send data in codes over optical fiber. It is similar to CDMA users that the data is coded in the forms of code and all are sent over single fiber at same wavelength. This basically improves the security of the data transmission.

There are various categories of codes as One Dimensional, Two Dimensional and Three Dimensional. In One Dimensional codes data is sent in wavelength domain. In case of Two Dimensional codes data is coded in two dimensions that is with wavelength, time is also included. In Three Dimensional codes third dimension of polarization is introduced.

2. TYPES OF AMPLIFIERS

A. Semiconductor optical amplifier (SOA)

Semiconductor optical amplifiers are amplifiers which use a semiconductor to provide the gain to the signal which is being given as a input to the optical amplifier. The semiconductor optical amplifier is of small size and being electrically pumped for it's operation.

The Semiconductor optical amplifier has some limitations like it has lower gain, moderate polarization dependence and high nonlinearity with fast transient time. The advantage of Semiconductor optical amplifier includes that different types of nonlinear operations can be conducted.

B. Erbium Doped Fiber Amplifier (EDFA)

The core of a optical fiber is doped with trivalent erbium ions and being efficiently pumped with a laser at a specific wavelength. The Erbium Doped Fiber Amplifier (EDFA) is the being deployed in fiber amplifier. EDFA is basically used in the third transmission window of silica-based optical fiber. Third window of transmission is made by the combination of two bands Conventional band (C-band) having wavelength ranges 1525 nm to 1565 nm, and Long band(L-band) having range 1570 nm to 1610 nm.

C. Fiber Raman Amplifier (FRA)

The signal is being amplified by Raman amplification. The main advantage of Raman amplification includes its ability to provide distributed amplification within the transmission fiber. FRA requires external pumping to perform the amplification operation, normally pump laser is used for pumping.

The paper has been divided in six parts. Part I gives an introduction to several types of amplifiers. Part II introduces the system design for OCDMA system. Schematic of the OCDMA system has been discussed in part III. Part IV includes result analysis. Conclusion is drawn in part V and future scope is mentioned in section VI.

3. SYSTEM DESIGN

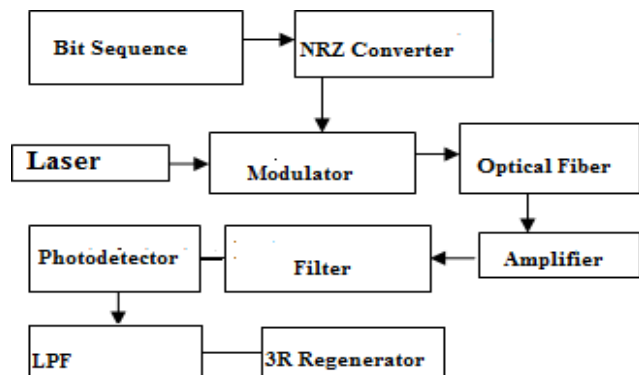


Fig. 1. Block Diagram of OCDMA System

A block diagram of an optical system used in paper to carry out results is shown in figure 1. It comprises of bit sequences (user defined and pseudo random both) followed by a non Return to Zero (NRZ) converter to convert the binary sequence to electrical pulse format. A multiplier is used to multiply the pseudo random and user defined sequences NRZ converted output before modulating with the laser. Output of NRZ converter is modulated with light source LASER. The data encoded from all users are multiplexed and passed over the optical fiber of defined length followed by a amplifier (either optical amplifier or hybrid amplifier) for amplification of information. Then amplified signal is given to fiber bragg grating (FBG) to filter the signal. The received signal is then converted back to electrical form with the help of photodetector, electrical signal is then passed to low pass filter to get the signal at desired wavelength and to filter out the unwanted wavelengths. Finally, it is given to the 3R regenerator for reamplification, reshaping and retiming of the signal. Design specifications considered in OCDMA system are mentioned in table I.

Table I: System Design parameters

Parameters	Values
Wavelength	1550 nm
Optical fiber length	1 km
Wavelength of Uniform FBG	1550 nm
FBG bandwidth	125 GHz
Optical Amplifier	SOA
Hybrid Amplifier	EDFA and FRA
Cut- off frequency of LPF	$0.75 * \text{bit rate Hz}$
Centre frequency of optical fiber	1550 nm
Fiber attenuation	0.2 dB/km

4. SCHEMATIC LAYOUT

The systems are designed for number of users with optical amplifier and a combination of EDFA and Raman amplifier i.e. hybrid amplifier. Simulations are carried out differently for systems with optical amplifier and hybrid amplifier. Schematic shown in figures 2(a) resembles the transmitter section and figure 2(b) & (c) resembles the receiver section with optical and hybrid amplifier for 5 users respectively. The LASER source is tuned at a single wavelength of 1550 nm. The modulate data is transmitted over 1 km optical fiber. At the receiver, optical amplifier is tuned at 193.4THz whereas EDFA and raman amplifiers are tuned at 193.4THz & 193.4THz respectively. FBG is set to same wavelength as that of the CW laser. Signal is now converted back to electrical domain by PIN photodetector and then passed through Bessel's low pass filter with cut off frequency of $0.75 * \text{Bit rate Hz}$ (bit rate is same as that of bit sequence at

the transmitter). The system performance using the above mentioned 2 types of amplifiers are studied for 2, 3,5,11 and 25 users at 1 Gbps data rate.

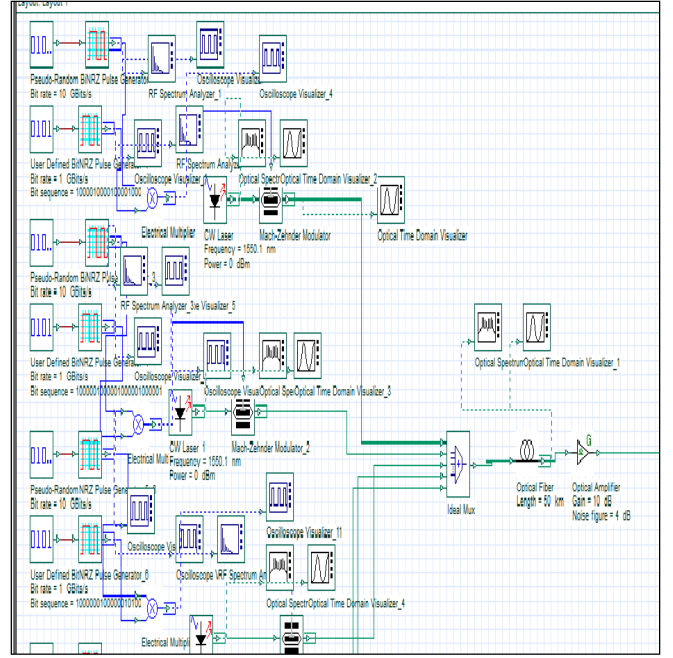


Fig. 2 (a). Transmitter Section of OCDMA System For 5 Users

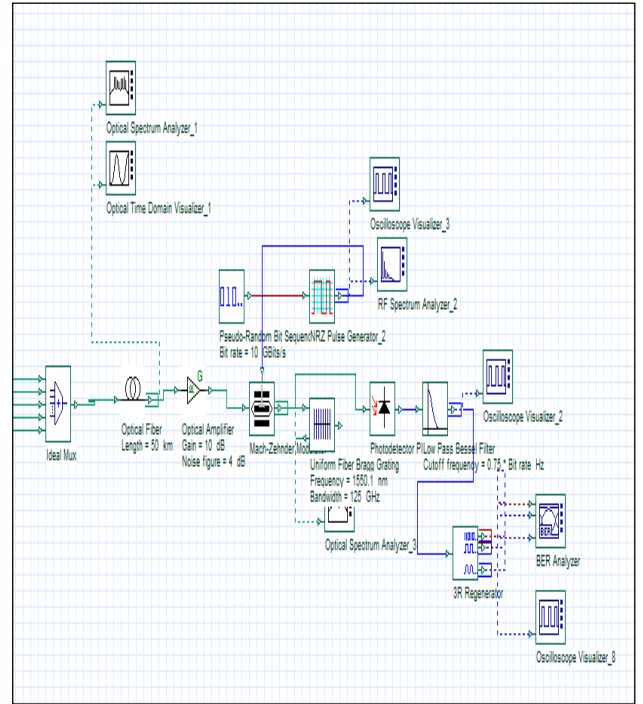


Fig. 2 (b). Receiver Section of OCDMA System for 5 Users with optical amplifier

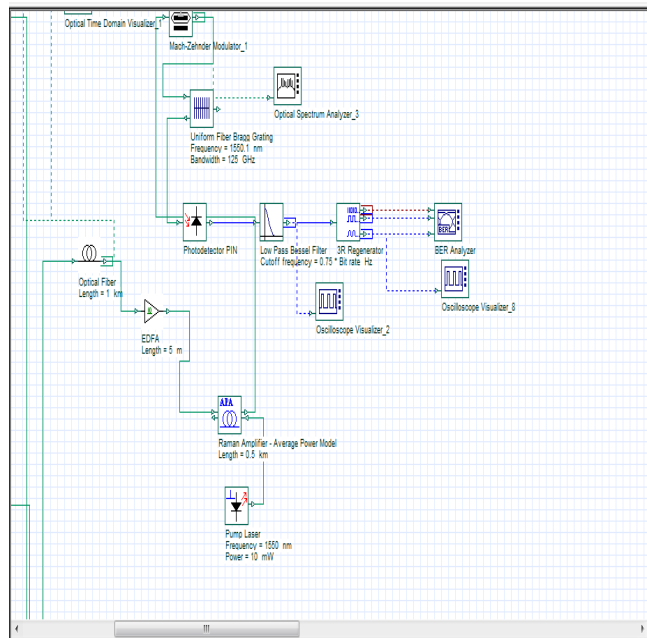


Fig. 2 (c) Receiver Section of OCDMA System For 5 Users with hybrid amplifier

5. RESULT ANALYSIS

Performance of the two systems discussed above, is evaluated on the basis of three different parameters: Bit error rate (BER), Q factor and eye height.

Systems are simulated and studied for 2,3,5, 11 and 25 users. An analysis of Q factor, BER performance and eye height of the two systems is shown graphically in the figures 3,4 and 5.

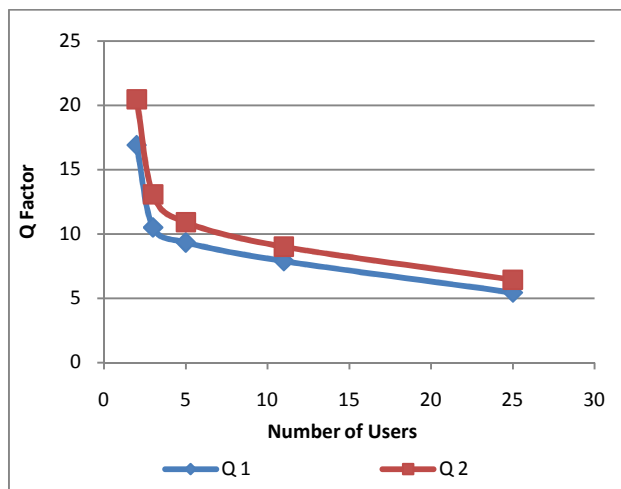


Fig. 3: Q Factor Analysis for two proposed systems.

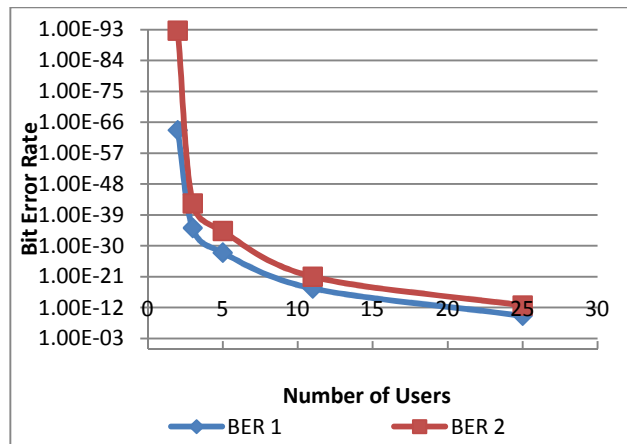


Fig. 4. BER Analysis for two proposed systems

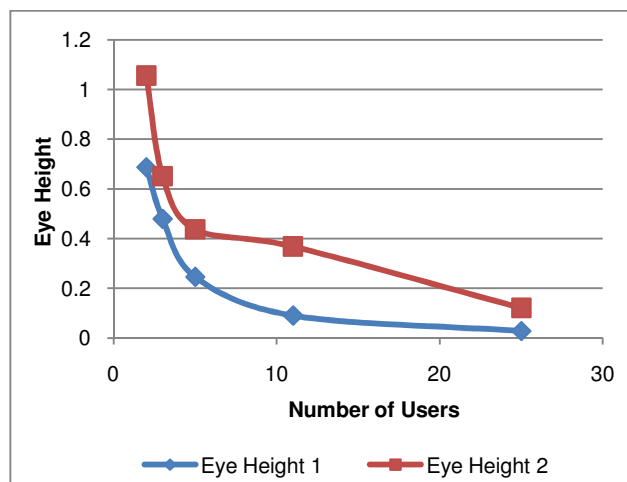


Fig. 5. Eye Height Analysis for two proposed systems

Comparative analysis of all three parameters shows that results are improved for OCDMA system using combination of EDFA and raman amplifier i.e. hybrid amplifier. Though parameters value decreases as number of users increases but as technology is enhancing number of users will increases day by day so a trade off will be required between parameters and number of users.

Table II :Q Factor for 2 Systems

No. Of users	Q 1	Q 2
2	16.9064	20.467
3	10.5123	13.0868
5	9.3474	10.9149
11	7.89901	9.01312
25	5.45864	6.44835

Table III: BER for 2 Systems

No. of users	BER 1	BER 2
2	1.93E-64	2.01E-93
3	5.96E-36	4.33E-43
5	9.58E-29	4.93E-35
11	2.58E-18	9.84E-22
25	2.39E-10	2.54E-13

Table IV: Eye Height for 2 Systems

No. of users	Eye Height 1	Eye Height 2
2	0.6871256	1.05582
3	0.4792785	0.651406
5	0.2457485	0.437465
11	0.0900763	0.368381
25	0.0279114	0.121171

Table II, III and IV shows the numerical values obtained for two systems. Values denoted by Q1, BER 1 and eye height 1 are for OCDMA system with optical amplifier and Q2, BER 2 and eye height 2 values are for hybrid amplifier (combination of EDFA and raman amplifier) system.

6. CONCLUSION

As seen from the graphs and tables performance of the

OCDMA system has increased due to introduction of hybrid amplifiers. Though cost will increase due to addition of EDFA and raman amplifier in OCDMA system but it is not a matter of thought that on side of increasing the performance cost can be overlooked.

7. FUTURE SCOPE

Hybrid amplifier constructed in paper is combination of EDFA and raman amplifier. Simulation can be done for more combination of amplifier using other types of amplifiers and even the number of users can be further increased. Even optical fiber length can be increased by implementing repeaters for lengthy fibers.

REFERENCES

- [1] G. Kaur and N. Gupta, "Design and performance analysis of 2.5Gbps OCDMA System by using newly constructed MPSC Code set for metropolitan area" in International Journal of engineering Research And industrial applications (IJERIA), vol. 2, no. II, pp. 245-257, 2009.
- [2] G. Kaur and N. Gupta, "Design and implementation of improved Superimposed Cyclic Optical Orthogonal Codes (SCOOC) based Optical Encoder/Decoder Structure for 1Gbps Optical CDMA System" in Journal of Engineering
- [3] Hongxi Yin, David J. Richardson, "Optical Code Division Multiple Access Communication Networks: Theory and Applications" e-book, Springer Publication, Tisungua University Press.
- [4] John M. Senior, "Optical Fibre Communication: Principles and Practice", Second Edition, Prentice Hall of India.

Review of ICI (Intercarrier Interference) Cancellation Techniques used in OFDM System

Ritika¹, Gurpriya Sandhu², Garima Saini³

^{1,2}M.E. Student, Department of Electronics and Communication Engineering

³Assistant Professor, Department of Electronics and Communication Engineering

National Institute of Technical Teachers Training & Research, Sector-26, Chandigarh, India

¹ritika876@gmail.com, ²gurpriasandhu@yahoo.com, ³garimasaini_18@rediffmail.com

Abstract: Orthogonal Frequency Division Multiplexing (OFDM) is a multicarrier modulation technique for the broadband wireless communication system. However, a main problem in OFDM is its vulnerability to frequency offset errors due to which the orthogonality is destroyed that result in Intercarrier Interference (ICI) in the OFDM symbol. ICI causes power leakage among subcarriers thus degrading the system performance. This paper aims to review the study of various techniques used for cancelling the Intercarrier interference (ICI). The ICI problem that occurs in OFDM system and countermeasure that can be taken for ICI cancellation are reviewed.

Keywords: Orthogonal Frequency Division Multiplexing (OFDM), Intercarrier Interference (ICI), multicarrier, frequency offset.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM), because of its resistance to multipath fading, has attracted increasing interest in recent years as a suitable modulation scheme for commercial high-speed broadband wireless communication systems. OFDM can provide large data rates with sufficient robustness to radio channel impairments. It is very easy to implement with the help of Fast Fourier Transform and Inverse Fast Fourier Transform for demodulation and modulation respectively [1].

It is a special case of multi-carrier modulation in which a large number of orthogonal, overlapping, narrow band sub-channels or subcarriers, transmitted in parallel, divide the available transmission bandwidth [2].

The separation of the subcarriers is theoretically minimal such that there is a very compact spectral utilization. These subcarriers have different frequencies and they are orthogonal to each other [3]. Since the bandwidth is narrower, each sub channel requires a longer symbol period. Due to the increased symbol duration, the ISI over each channel is reduced.

However, a major problem in OFDM is its vulnerability to frequency offset errors between the transmitted and received signals, which may be caused by Doppler shift in the channel or by the difference between the transmitter and receiver local oscillator frequencies [4]. In such situations, the orthogonality of the carriers is no longer maintained, which results in Intercarrier Interference (ICI). ICI results from the other sub-channels in the same data block of the same user. ICI problem would become more complicated when the multipath fading is present [5]. If ICI is not properly compensated it results in power leakage among the subcarriers, thus degrading the system performance.

The objective of this paper is to done literature review of ICI cancellation in OFDM system.

Organization: This paper is organized as follows: In section II the Literature is reviewed; In section III conclusion is given with future scope. In last the references are given.

2. LITERATURE REVIEW

The purpose of this literature review is to study the literature of emergence of ICI problem in OFDM system and review of various methods used to mitigate the ICI in OFDM systems which are described by various authors in certain papers.

A) System Description

The block diagram of standard OFDM system is given in Figure 1. In an OFDM system, the input data stream is converted into N parallel data streams each with symbol period T_s through a serial-to-parallel Port. When the parallel symbol streams are generated, each stream would be modulated and carried over at different center frequencies. The sub-carriers are spaced by $1/NT_s$ in frequency, thus they are orthogonal over the interval $(0, T_s)$. Then, the N symbols are mapped to bins of an inverse fast Fourier transform (IFFT). These IFFT [6] bins correspond to

the orthogonal sub-carriers in the OFDM symbol. Therefore, the OFDM symbol can be expressed as

$$x(n) = \frac{1}{N} \sum_{m=0}^{N-1} X_m e^{j2\pi mn/N} \quad (1)$$

Where the X_m 's are the base band symbols on each sub-carrier. The digital-to-analog (D/A) converter then creates an analog time-domain signal which is transmitted through the channel.

At the receiver, the signal is converted back to a discrete N point sequence $y(n)$, corresponding to each sub-carrier. This discrete signal is demodulated using an N-point Fast Fourier Transform (FFT) operation at the receiver.

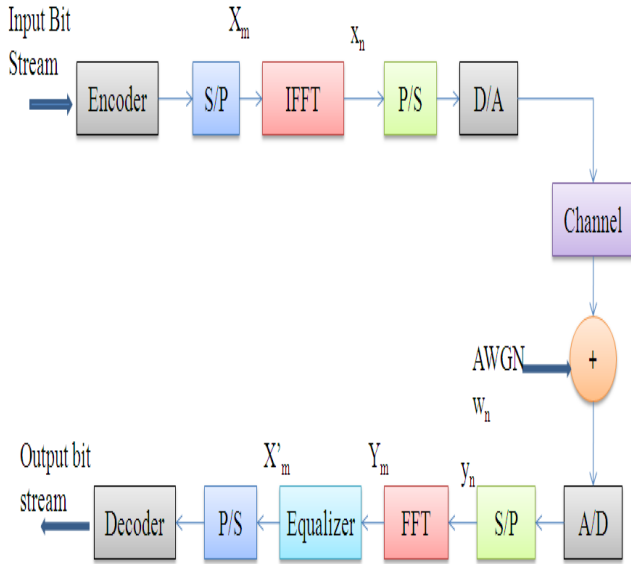


Fig. 1. OFDM System Model

The demodulated symbol stream is given by:

$$Y(m) = \sum_{n=0}^{N-1} y(n) e^{-j2\pi mn/N} + w(m) \quad (2)$$

where $w(m)$ corresponds to the FFT of the samples of $w(n)$, which is the Additive White Gaussian Noise (AWGN) introduced in the channel.

B) Analysis of inter-carrier interference

The main disadvantage of OFDM, however, is its susceptibility to small differences in frequency at the

transmitter and receiver, normally referred to as frequency offset. This frequency offset can be caused by Doppler shift due to relative motion between the transmitter and receiver, or by differences between the frequencies of the local oscillators at the transmitter and receiver. The frequency offset is modeled as a multiplicative factor introduced in the channel, as shown in Figure 2.

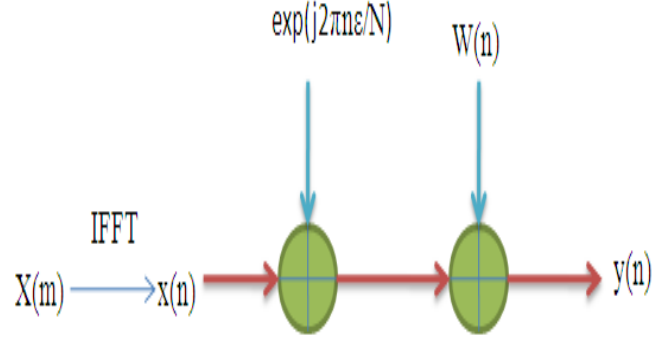


Fig. 2. Frequency Offset Model

The received signal is given by,

$$Y(n) = x(n) e^{j2\pi n\epsilon/N} + w(n) \quad (3)$$

where ϵ is the normalized frequency offset, and is given by $\Delta f N T_s$. Δf is the frequency difference between the transmitted and received carrier frequencies and T_s is the subcarrier symbol period. $W(n)$ is the AWGN introduced in the channel.

The effect of this frequency offset on the received symbol stream can be understood by Considering the received symbol $Y(k)$ on the k sub-carrier

$$Y(k) = X(k) S(0) + \sum_{l=0, l \neq k}^{N-1} X(l) S(l-k) + n_k \quad (4)$$

$k = 0, 1, \dots, N-1$

where N is the total number of subcarriers, $X(k)$ is the transmitted symbol for the k subcarrier, n_k is the FFT of $w(n)$, and $S(l-k)$ are the complex coefficients for the ICI components in the received signal. The ICI components are the interfering signals transmitted on sub-carriers. The complex coefficients are given by,

$$S(l-k) = \frac{\sin(\pi(l+\epsilon-k))}{N \sin(\pi(l+\epsilon-k)/N)} \exp(j\pi(1-\frac{1}{N})(l+\epsilon-k)) \quad (5)$$

C. ICI cancellation schemes

Table I. Summarized Table of The Review

ICI cancellation scheme	Author	Advantages	Disadvantages	Summary
Frequency Domain Correlative Coding (7)	Yu Zhang, Huaping Liu	1.This coding is used to increase CIR and improves the BER performance.	1. This technique cannot be used in frequency selective fading channels.	Derive expression for CIR to show the impact of time-selective fading and demonstrate the effectiveness of correlative coding in mitigating ICI in MIMO-OFDM systems. For simulation they consider a system with two transmit antennas and two receive antennas which employs BPSK modulation and adopt the "SUI-5" channel model.
Self Cancellation technique (8)	Yuping Zhao and Sven-Gustav Häggman	1.Leads to high CIR. 2.More power efficient. 3.Best results for small freq. offset and binary alphabet size	1.Reduces Bandwidth efficiency by half. 2.Performance not good for high frequency offset.	Modulate the input data symbol onto a group of subcarriers with predefined coefficients such that the generated ICI signals within that group cancel each other.
Maximum Likelihood(9)	P.H. Moose	1.Lower BER at low and high frequency offset	1.Complex implementation	The frequency offset is first statistically estimated using a maximum likelihood algorithm and then cancelled at the receiver.
Time-Domain Equalization (10)	R.Kumar, Malarvizhi	1.it provides better BER performance.	1.Not perform well for frequency selective fading channels.	Proposed a window function which creates a correlation between two adjacent subcarriers and gives a higher signal to ICI ratio than standard OFDM.
Carrier conjugate(11)	Hen-Geul Yeh, Yuan-Kwei Chang, Babak Hassibi	1.High signal to interference power ratio (SIR) in the presence of small frequency offsets. 2.Better bit error rate (BER) performance in both additive white Gaussian noise (AWGN) and fading channels.	1.Reduces Bandwidth efficiency.	Uses a two path Algorithm: The first path uses the regular OFDM algorithm. The second path uses the conjugate transmission of the first path. The combination of both paths forms a conjugate ICI cancellation scheme at the receiver.
Repeated Correlative Coding(12)	V.K. Dwivedi, G.Singh	1. Provides better CIR as compared to self-cancellation and correlative coding.	1.Not provide satisfactory performance for 16-QAM Modulation.	This scheme combines two methods, which are the coding of adjacent subcarriers with antipodal of the same data symbol (ICI self-cancellation) and correlative coding. Derives the expression for CIR(carrier to interference ratio) and compared the simulated result with correlative coding and self-cancellation technique.

3. CONCLUSION

In this paper, the ICI cancellation schemes like Frequency domain correlative coding, Self-cancellation, Maximum Likelihood, Time domain equalization, CC and Repeated Correlative Coding schemes have been reviewed in terms of the Carrier-to-Interference ratio (CIR) and the bit error rate (BER) performance. Inter-carrier interference (ICI) which results from the frequency offset degrades the performance

of the OFDM system. Every scheme has its own advantages and disadvantages.

The choice of which method to employ depends on the specific application. For example, self cancellation does not require very complex hardware or software for implementation. However, it is not bandwidth efficient as there is a redundancy of 2 for each carrier. The ML method also introduces the same level of redundancy but provides

better BER performance, since it accurately estimates the frequency offset. Its implementation is more complex than the SC method. On the other hand, CC provides better results for both low and high frequency offsets and for flat-fading and multipath fading but it reduces bandwidth efficiency. Also Repeated Correlative Coding schemes provides better CIR and BER as compared to Correlative Coding and Self-ICI Cancellation Techniques. The performance of all schemes is compared by considering AWGN channel.

4. FUTURE SCOPE

Further work can be done by performing simulations to investigate the performance of these ICI cancellation schemes in multipath fading channels without perfect channel information at the receiver.

REFERENCES

- [1] Ramjee Prasad, "OFDM for wireless communication system", Artech House, 2004.
- [2] S.Weinstein and P.Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform", IEEE Trans. Commun., Vol. 19, pp. 628-634, Oct. 1971.
- [3] L.J. Climini, "Analysis and Simulation of a digital mobile channel using orthogonal frequency division multiplexing", IEEE Trans. on Communications, No. 7, July 1985.
- [4] M. Russell, G.L. Stuber, " Interchannel interference analysis of OFDM in a mobile environment", Vehicular Technology Conference, 45th IEEE Conference, Vol. 2, pp. 820-824, July 1995.
- [5] X.Cai, G.B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM", IEEE Trans. on Communications, Vol. 51, pp. 2047-2056, No. 12, December 2003.
- [6] William H. Tranter, K. Sam Shanmugam, Theodore S.Rappaport, "Principles of communication system simulation with wireless application" Pearson Education, 2004.
- [7] Yu Zhang, Huaping Liu, "Frequency-Domain correlative coding for MIMO-OFDM systems over fast fading channels", IEEE Communications Letters, Vol. 10, No. 5, May 2006.
- [8] Y. Zhao, S.G. Haggman, "Inter-carrier interference self-cancellation scheme for OFDM mobile communication systems", IEEE Trans. Commun., Vol. 49, No. 7, pp. 1185-1191, July 2001.
- [9] P.H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction", IEEE Trans. Commun., Vol. 42, No. 10, October 1994.
- [10] Kumar, S. Malarvirzhi, S. Jayashri, "Time-Domain Equalisation technique for intercarrier interference suppression in OFDM systems", Information Technology Journal, No. 7, pp. 149-154, 2008.
- [11] Hen-Guel Yeh, Yuan-Kwei Chang, Babak Hassibi, "A scheme for cancelling intercarrier interference using conjugate transmission in multicarrier communication systems", IEEE Trans. Commun., Vol. 6, No. 1, January 2007.
- [12] V.K. Dwivedi, G.Singh, "Repeated Correlative coding scheme for mitigation of intercarrier interference in an orthogonal frequency division multiplexing system", IET Commun. Vol. 6, No. 6, pp. 599-603, June 2012.

A Survey on Fiber Optic Sensors

Pinky Khundrakpam¹, Priyanka Sharma²

^{1,2}Amity Institute of Telecom Technology and Management, Noida, India
¹pinks.kh@gmail.com, ²priyanka433190@gmail.com

Abstract: Photonic sensors, signal processors and communication technologies have come up as alternatives of electronics. Fiber optic sensors have progressed to such an extent that they can be readily made use of in many applications. The paper presents different types of fiber optic sensors and components used in optical fiber sensing.

1. INTRODUCTION

The advent of lasers and semiconductor optoelectronics, revolutionise modern optics. Rapid developments in fiber optics gave an additional impetus to this. A wide range of fiber optic sensors are fast reaching the market place. Freedom from electromagnetic interference and unprecedented speed and bandwidth are prime reasons for all the success in photonics. Impressive though the developments in photonics till date is, it is yet to reach its full potential especially in terms of the market exploitation. One of the exciting fields wherein photonics is expected to play a significant role is smart structures and intelligent systems of interest in engineering. This is where the real challenge lies even in the case of fiber optic sensors and photonic communication and control schemes. In smart structure applications, composite materials, fiber optic sensing and telemetry systems, piezoelectric actuators and microprocessor based control schemes seem to offer the best advantages as of now.

2. FIBRE OPTIC SENSORS

The technology and applications of optical fibers have progressed very rapidly in recent years. Optical fiber, being a physical medium, is subjected to perturbation of one kind or the other at all times. It therefore experiences geometrical (size, shape) and optical (refractive index, mode conversion) changes to a larger or lesser extent depending upon the nature and the magnitude of the perturbation.

In communication applications one tries to minimize such effects so that signal transmission and reception is reliable. On the other hand in fiber optic sensing, the response to external influence is deliberately enhanced so that the resulting change in optical radiation can be used as a measure of the external perturbation. In communication, the signal passing through a fiber is already modulated, while in sensing, the fiber acts as a modulator. It also serves as a transducer and converts measurands like temperature, stress,

strain, rotation or electric and magnetic currents into a corresponding change in the optical radiation. Since light is characterized by amplitude (intensity), phase, frequency and polarization, any one or more of these parameters may undergo a change. The usefulness of the fiber optic sensor therefore depends upon the magnitude of this change and our ability to measure and quantify the same reliably and accurately.

The advantages of fiber optic sensors are freedom from EMI, wide bandwidth, compactness, geometric versatility and economy. In general, FOS is characterized by high sensitivity when compared to other types of sensors. It is also passive in nature due to the dielectric construction. Specially prepared fibers can withstand high temperature and other harsh environments. In telemetry and remote sensing applications it is possible to use a segment of the fiber as a sensor gauge while a long length of the same or another fiber can convey the sensed information to a remote station.

Deployment of distributed and array sensors covering extensive structures and geographical locations is also feasible. Many signal processing devices (splitter, combiner, multiplexer, filter, delay line etc.) can also be made of fiber elements thus enabling the realization of an all-fiber measuring system. Recently photonic circuits (Integrated Optics) has been proposed as a single chip optical device or signal processing element which enables miniaturization, batch production, economy and enhanced capabilities.

There are a variety of fiber optic sensors. These can be classified as follows.

A) Based on the modulation and demodulation process a sensor can be called as an intensity (amplitude), a phase, a frequency, or a polarization sensor. Since detection of phase or frequency in optics calls for interferometric techniques, the latter are also termed as interferometric sensors. From a detection point of view the interferometric technique implies heterodyne detection/coherent detection. On the other hand intensity sensors are basically incoherent in nature. Intensity or incoherent sensors are simple in construction, while coherent detection (interferometric) sensors are more complex in design but offer better sensitivity and resolution.

B) Fiber optic sensors can also be classified on the basis of their application: physical sensors (e.g. measurement of temperature, stress, etc.); chemical sensors (e.g. measurement of pH content, gas analysis, spectroscopic studies, etc.); bio-medical sensors (inserted via catheters or endoscopes which measure blood flow, glucose content and so on). Both the intensity types and the interferometric types of sensors can be considered in any of the above applications.

C) Extrinsic or intrinsic sensors is another classification scheme. In the former, sensing takes place in a region outside of the fiber and the fiber essentially serves as a conduit for the to-and-fro transmission of light to the sensing region efficiently and in a desired form. On the other hand, in an intrinsic sensor one or more of the physical properties of the fiber undergo a change as mentioned in A) above.

3. BASIC COMPONENTS

A fiber optic sensor in general will consist of a source of light, a length of sensing (and transmission) fiber, a photodetector, demodulator, processing and display optics and the required electronics.

1 **Optical fibers:** These are thin, long cylindrical structures which support light propagation through total internal reflection. An optical fiber consists of an inner core and an outer cladding typically made of silica glass, although, other materials like plastics are some times used. Three types of fibers are in common use in FOS. The multimode (MM) fiber consists of a core region whose diameter ($\sim 50 \mu m$) is a large multiple of the optical wavelength. The index profile of the core is either uniform (step-index) or graded (eg., parabolic). Plastic fibers have a step index profile and a core size of about 1mm. The microbend type or the evanescent type intensity sensors use MM fibers. MM fiber has the advantage that it can couple large amount of light and is easy to handle, both the advantages arising from its large core size.

Single mode (SM) fiber is designed such that all the higher order waveguide modes are cut-off by a proper choice of the waveguide parameters as given below.

$$V = \frac{2\pi a}{\lambda} \sqrt{n_1^2 - n_2^2}$$

where, λ is the wavelength, a is the core radius, and n_1 and n_2 are the core and cladding refractive indices, respectively. When $V < 2.405$ single mode condition is ensured. SM fiber is an essential requirement for interferometric sensors. Due to the small core size ($\sim 4 \mu m$) alignment becomes a critical factor.

The SM fiber mentioned above single mode in that two modes with degenerate polarization states can propagate in the fiber. This can lead to signal interference and noise in the measurement. The degeneracy can be removed and a single mode polarization preserving fiber can be is not truly obtained by the use of an elliptical core fiber of very small size or with built in stress. In either case, light launched along the major axis of the fiber is preserved in its state of polarization. It is also possible to make a polarizing fiber in which only one state of polarization is propagated. Polarimetric sensors make use of polarization preserving fibers. Thus, multimode fiber, single mode fiber and polarization preserving fiber are the three classes of fibers which are used in the intensity type, the interferometric type and the polarimetric type of sensors, respectively.

2 **Sources:** In FOS semiconductor based light sources offer the best advantages in terms of size, cost, power consumption and reliability. Light emitting diodes (LEDs) and laser diodes (LDs) are the right type of sources for FOS although in laboratory experiments the He-Ne laser is frequently used. Features of LED include very low coherence length, broad spectral width, low sensitivity to back reflected light and high reliability. They are useful in intensity type of sensors only.

LDs on the other hand exhibit high coherence, narrow linewidth and high optical output power, all of which are essential in interferometric sensors. Single mode diode lasers are made using distributed feedback or external cavity schemes. High performance Mach-Zehnder and Fabry-Perot type sensors need single mode lasers. LDs in general are susceptible to reflected (feedback) light and temperature changes. They are also less reliable and more expensive. Coupling of light from source to fiber is an important aspect and may call for special optical devices. Use of pigtailed source can alleviate this problem but such devices cost more. Fiber lasers and amplifiers are fast becoming commercial products and may play an important role in future FO sensors.

3 **Detectors:** Semiconductor photodiodes (PDs) and avalanche photodiodes (APDs) are the most suitable detectors in FOS. APD can sense low light levels due to the inherent gain because of avalanche multiplication, but need large supply voltage typically about 100 V. The various noise mechanisms associated with the detector and electronic circuitry limit the ultimate detection capability. Thermal and shot noise are two main noise sources and need to be minimized for good sensor performance. Detector response varies as a function of wavelength. Silicon PD is good for visible and near IR wavelengths. Generally there is no bandwidth limitation due to the detector as such, although the associated electronic circuits can pose some limitation.

4. CONCLUSIONS

The most obvious example of a fiber optic sensor succeeding in this arena is the fiber optic gyro, which is displacing both mechanical and ring laser gyros for medium-accuracy devices. Fiber optic sensors are being embedded into or attached to materials (1) during the manufacturing process to enhance process control systems, (2) to augment non destructive evaluation once parts have been made, (3) to form health and damage assessment systems once parts have been assembled into structures, and (4) to enhance control systems. Fiber optic sensors can be embedded in a panel and multiplexed to minimize the number of leads. The signals from the panel are fed back to an optical=electronic processor for decoding. The information is formatted and transmitted to a control system that could be augmenting performance or assessing health. The control system would

then act, via a fiber optic link, to modify the structure in response to the environmental effect

REFERENCES

- [1] Dakin and B.Culshaw (Eds.), "Optical Fiber Sensors, Principles and Components", Vol. I & II, Artech House, Boston 1988.
- [2] Eric Udd (ed.), "Fiber Optic Sensors, An introduction for Engineers and Scientists", John Wiley & Sons, Inc. New York, 1991.
- [3] D. A. Krohn, Fiber Optic Sensors: Fundamentals and Applications, Instrument Society of America, Research Triangle Park, NC, 1988.
- [4] E. Udd and P. M. Turek, Single mode fiber optic vibration sensor, Proc. SPIE, 566, p. 135, 1985.
- [5] W. W. Morey, Distributed fiber grating sensors, Proc. 7th Optical Fiber Sensor Conf., IREE Australia, Sydney, p. 285, 1990

Holography: Real Cyber World

Aprajita Sharma¹, Ashish Gupta²

^{1,2}*Electronics and Telecommunication*

Amity Institute of Telecom Technology and Management, Noida, India

¹*aprajita.sharma19@gmail.com*, ²*ashish.research@yahoo.com*

Abstract: Recent advances in both the computation and display of holographic images have enabled several firsts. Interactive display of images is now possible using the bipolar intensity computation method and a fast look-up table approach to fringe pattern generation. Full-color images have been generated by computing and displaying three color component images (red, green, and blue). Using parallelism to scale up the first generation system, images as large as 80 mm in all three dimensions have been displayed. The combination of multi-channel acousto-optic modulators and fast horizontal scanning continue to provide the basis of an effective real-time holographic display system.

1. INTRODUCTION

Three-dimensional video displays that can generate ghost-like optical duplicates of 3-D objects and scenes have been depicted in science-fiction movies as futuristic means of visual media tools; such display devices always attracted public interest [1]. Holography in comparison to 3D-stereo overcomes the problem of the depth-cue mismatch between depth-focus and convergence. This so called accommodation-convergence mismatch leads to fatigue or headache, even a short loss of orientation may occur, so with 3D-stereo, only small depth-ranges must be realized and the time to consume 3D-stereo without a break should also be very limited. Holography in contrast is like natural 3D-viewing, which allows very large depth ranges, there are no negative effects, because the eyes can both focus and converge on the object seen. When looking at a hologram, the object focused looks sharp while other objects in different distances will look blurry like it is in real life.

One immediate question is whether such a display is possible; and a quick answer is B Yes, it is. Noting that B seeing is a purely optical interaction, and what we (or any other observer, including living organisms and machines) see is only due to the light that enters through our pupils, the design target for such a display is simple to state: if we can record the volume filling time-varying light field in a 3-D scene, with all its needed physical properties, and then regenerate the same light field somehow at another place, maybe at another time, the observer will not be able to distinguish the original scene from its duplicate since the received light will be the same, and therefore, any visual perception will also be the same. Then the natural question is whether we can record the light with all its relevant physical

properties, and then regenerate it. The classical video camera is also a light recorder. However, not all necessary physical properties of light for the purpose outlined above can be recorded by a video camera; indeed, what is recorded by a video camera is just the focused intensity patterns (one for each basic color) over a planar sensing device. What is needed to be recorded instead is indeed much more complicated: we also need the directional decomposition of incoming light as well. Briefly, and in an idealized sense, we can say that we need to record the light field distribution. The term light field distribution is usually associated with ray optics concepts, and therefore, can be a valid optics model only in limited cases. If it can be recorded, we then need physical devices that can also regenerate (replay) the recorded light field. Prototypes for light field recording and rendering devices are reported in the literature [2].

Integral imaging gets close to a light field imaging device in the limit under some mathematical idealizations; however such limiting cases are not physically possible [3]. A better optical model than the ray optics is the wave optics. The propagation of light in a volume is modeled as a scalar wave field; the optical information due to a 3-D scene is carried by this wave field. Therefore, if such a wave field can be recorded and replayed, we achieve visual duplication of 3-D scenes; this is holography [4]. Scalar wave model is usually satisfactory, and more accurate models of light are rarely needed, if any, for 3-D imaging and display purposes. Therefore, the term holography refers to recording and replaying optical wave fields. In a more restrictive usage, holography refers only to a specific form of such recording where interference of the desired wave field with a reference wave (sometimes self-referencing is employed, as in in-line holography) is formed and recorded; we prefer the broader usage as stated above. Indeed, the usage of the term may even be further broadened to include all kinds of physical duplication of light, and therefore, may also cover integral imaging, in a sense [1].

Here in this paper, our focus is on the display of holograms. We focus only on dynamic displays for video. Still holographic display technology has been well developed since 1960s, whereas dynamic display technology is still in its infancy, and therefore, a current research topic. We further restrict our focus to pixelated display devices that can be driven digitally. Such displays are usually called digital

electroholographic displays since they are driven electronically. An overview of some research results in this field, together with current research interests and achievements, will be presented in Section II. Section III presents an analysis to understand the effects of different parameters to the holographic reconstruction quality; the analysis then leads the specifications of a satisfactory quality digital dynamic holographic display. the resolution is significantly lower compared to thick holograms. Moreover, pixelated structures bring some additional problems. Pixel period determines the maximum frequency that can be represented when digital-to-analog conversion is conducted in the Shannon sense, and this in turn determines the maximum diffraction angle as outlined in Section III-A2. Unfortunately, the pixel periods are not currently small enough to support sufficiently large viewing angles. Problems associated with pixelated electroholographic display are known [11].

Since liquid crystal spatial light modulators (SLMs) are currently the primary choice for digital holographic displays, it is quite relevant to briefly mention current capabilities of such devices. Bauchert et al. [12] reported the desirable features of liquid crystal SLMs. These features can be summarized as higher number of pixels, smaller pixel period, better optical efficiency, and faster operation. There are various SLMs such as liquid-crystal-based devices (liquid-crystal devices and liquid crystal on silicon devices), mirror based devices (digital micromirror devices) and solid crystal devices (acousto-optical devices). The acousto-optical modulators (AOMs) are mostly used in 1-D applications. The digital micromirror devices are usually for binary modulation and they may result in additional noise due to vibration of micromirrors. The liquid-crystal-based light modulators are more commonly used in electroholographic applications. Michalkiewicz et al. presented the progress in liquid crystal on silicon (LCoS) SLMs and their applications [13]. Ohmura et al. proposed a method to increase the viewing angle using such SLMs [14]. In their proposed system, they used a single SLM that was driven by a mirror module. As a consequence of this method the resolution along the horizontal direction increases.

Therefore, the horizontal diffraction angle also increases; and thus the viewing angle is improved. Liquid-crystal-based SLMs are classified into various types such as complex amplitude, amplitude-only, phase-only, transmissive- and reflective-type SLMs, and so on. Ability to support complex functions at the display is highly desirable since diffraction fields are represented as complex valued fields where both the amplitude and the phase are needed. An ideal SLM pixel should modulate both the amplitude and the phase of the incident light. However, it is difficult to manufacture the complex amplitude-type SLMs based on current technology. Phase-only SLMs may be the next best solutions for electroholography because they have

several advantages over amplitude-only SLMs such as suppressed zeroth-order and high-diffraction efficiency, which can theoretically reach 100%. Amplitude-only SLMs can also be used for electroholography. However, problems associated with strong undesired diffraction orders are more severe compared to the phase-only case. A research group from Barcelona University, Barcelona, Spain, combined two SLMs to display full complex Fresnel holograms [15].

They used one SLM for the amplitude and the other one for the phase. They also investigated the quality of the reconstructions using real-only, imaginary-only, amplitude-only, and phase-only holograms. Schwerdtner et al. reported a novel hologram technology, which they called tracked viewing window (TVW) [16]. By this approach they only calculate a small portion of a hologram, which then reconstructs a narrow angle light that falls onto the tracked pupils of the observer. They demonstrated that thin film transistor (TFT) monitors can then be used as SLMs to build holographic displays. Another electroholographic display technique was presented by Hahn et al. [17].

In their research, they used curved array of SLMs to increase the field of view. Spatial Imaging Group at the Massachusetts Institute of Technology (MIT, Cambridge, MA) developed a series of holographic display systems named Mark-I, Mark-II, and Mark-III [18]. Mark-I and Mark-II use acousto-optical modulators, whereas Mark-III uses guided-wave optical scanners. All three can render 3-D objects at video rates. A company developed another holographic display system [19]. The system uses active tiling where an electrically addressed SLM (EASLM) projects tiles of a big hologram onto an optically addressed SLM (OASLM). With the help of the setup, more than 100 megapixels holograms can be displayed. Another system, so-called Horn (Holographic Reconstruction), was presented by a group in Chiba University, Chiba, Japan [20]. Field-programmable gate arrays (FPGAs) were used in the developed holographic display system to achieve video frame rates. Another group from Japan also demonstrated a holographic display system [21]; the system at the National Institute of Information and Communications Technologies (NICT, Tokyo, Japan) captures the 3-D scene by an integral imaging camera. The digital holograms of the captured scene is calculated and displayed in real time.

2. STATE OF THE ART IN DYNAMIC HOLOGRAPHIC DISPLAYS

A. Overview Even though we focus on dynamic holographic displays in this paper, we feel that it is appropriate to start with a brief history of holography in general. Gabor (1900–1979) invented the holography to reduce the aberrations in electron microscopy [5]. However, due to low quality of obtained images holography did not become popular until early 1960s. After the developments in laser technologies,

Leith and Upatnieks [6] developed the off-axis holography. In the meantime, Denisyuk invented the volume holography by bringing the work of Lippmann to holography [7]. Still holography has been significantly developed since then, and many excellent monochromatic and color holograms have been made. The first computer generated hologram was introduced by Lohmann and Paris in 1967 [8]. In the same year, Goodman and Lawrence brought forward the idea of the digital HOLOGRAPHY [9].

Then, in 1980, The Fundamental Theory of digital holography was introduced by Yaroslavskii and Merzlyakov [10]. We use the term digital holography in a broader sense to include all sorts of digital techniques to compute wave propagation, diffraction, and interference, as well as, digital capture and digital display of holograms. Conventional thick holograms on photographic plates can provide high resolution and full parallax. However, dynamic displays for holographic video are still far from providing satisfactory results.

3. SCHEMATIC FOR AN ELECTRO-HOLOGRAPHIC DISPLAY

The physical parameters of the human visual system, such as the field of view, spatial resolution, visible spectrum, and so on, are naturally limited. Therefore, the spatial bandwidth of an electroholographic display device may also be limited based on the corresponding limitations of the human visual system. A simple schematic for an electroholographic display system is shown in Fig.1 A plane wave illuminates the hologram. This beam is diffracted by the SLM, with an area of S_H , towards multiple directions within an angle $2\theta_{\max}$. The distance between the observer and the hologram is denoted as D_H and the angular field of view of the observer is $2\phi_v$ [22]. This schematic will be used to find the related specifications.

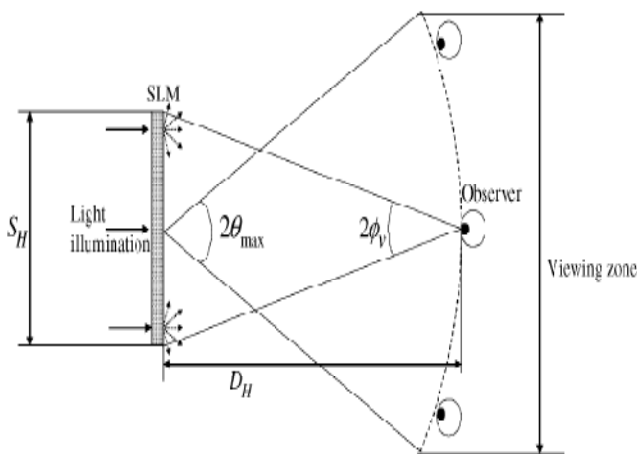


Fig. 1. Simple schematic for an electroholographic display

4. CONCLUSION

Digital holographic video displays are strong candidates for rendering ghost-like but true 3-D motion images. Interest in this technology is increasing among the research community. Since the holograms are quite robust to quantization errors, and since frame refresh rates are satisfactory for continuous perception, the focus of research is rather on designing digital holographic display sets, which can effectively support more freedom in lateral and rotational motion of the observer while providing satisfactory quality 3-D images.

REFERENCES

- [1] L. Onural, Television in 3-D: What are the prospects?'' Proc. IEEE, vol. 95, no. 6, pp. 1143–1145, Jun. 2007.
- [2] M. Levoy, Light fields and computational imaging, [Computer, pp. 46–55, Aug. 2006.
- [3] E. Sahin and L. Onural, A comparative study of light field representation and integral imaging, [Imag. Sci. J., vol. 58, pp. 28–31, 2010.
- [4] P. Hariharan, Optical Holography: Principles, Techniques and Applications, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [5] D. Gabor, A new microscopic principle, [Nature, vol. 161, pp. 777–778, 1948.
- [6] E. N. Leith and J. Upatnieks, Wavefront reconstruction with diffused illumination and three-dimensional objects, [J. Opt. Soc. Amer., vol. 54, pp. 1295–1301, 1964
- [7] Y. N. Denisyuk, BPhotographic reconstruction of the optical properties of an object in its own scattered field, [Sov. Phys. Dokl., vol. 7, p. 543, 1962.
- [8] A. W. Lohmann and D. Paris, Binary Fraunhofer holograms generated by computer, [Appl. Opt., vol. 6, pp. 1739–1748, 1967.
- [9] J. W. Goodman and R. W. Lawrence, Digital image formation from electronically detected holograms, [Appl. Phys. Lett., vol. 11, pp. 77–79, 1967
- [10] L. P. Yaroslavskii and N. S. Merzlyakov, Methods of Digital Holography. New York: Consultants Bureau, 1980.
- [11] F. Yaraz, M. Kovachev, R. Ilieva, M. Agour, and L. Onural, Holographic reconstructions using phase-only spatial light modulators, [in Proc. 3DTV Conf., The True Vision V Capture, Transmission and Display of 3D Video, 2008,
- [12] K. Bauchert, S. Serati, and A. Furman, Advances in liquid crystal spatial light modulators, [Proc. SPIEVOpt. Pattern Recognit. XIII, vol. 4734, pp. 35–43, 2002
- [13] A. Michalkiewicz, M. Kujawinskaa, T. Kozackia, X. Wangb, and P. J. Bosb, Holographic three-dimensional displays with liquid crystal on silicon spatial light modulator, [Proc. SPIE V Interferometry XII: Tech. Anal., vol. 5531, pp. 85–94, 2004.
- [14] N. Ohmura, H. Kang, T. Yamaguchi, and H. Yoshikawa, A method to increase the hologram viewing angle by the beam reconfiguration, [Proc. SPIE V Practical Holography XXII: Mater. Appl., vol. 6912, 69120O, 2008.
- [15] R. Tudela, I. Labastida, E. Marti-Badosa, S. Vallmitjana, I. Juvells, and A. Carnicer, A simple method for displaying

- fresnel holograms on liquid crystal panels, [Opt. Commun., vol. 214, pp. 107–114, 2002
- [16] A. Schwerdtner, R. Haussler, and N. Leister, Large holographic displays for real-time applications, [Proc. SPIE, vol. 6912, no. 1, 69120T, 2008.
- [17] J. Hahn, H. Kim, Y. Lim, G. Park, and B. Lee, Wide viewing angle dynamic holographic stereogram with a curved array of spatial light modulators, [Opt. Exp., vol. 16, no. 16, pp. 12372–12386, 2008.
- [18] P. S. Hilaire, S. A. Benton, M. Lucente, and P. M. Hubel, Color images with the MIT holographic video display, [Proc. SPIE Practical Holography VI, vol. 1667, pp. 73–84, 1992
- [19] M. Stanley, M. A. Smith, A. P. Smith, P. J. Watson, S. D. Coomber, C. D. Cameron, C. W. Slinger, and A. D. Wood, 3D electronic holography display system using a 100 megapixel spatial light modulator, [Proc. SPIE, vol. 5249, pp. 297–308, 2004
- [20] T. Shimobaba, S. Hishinuma, and T. Ito, Special-purpose computer for holography HORN-4 with recurrence algorithm, [Comput. Phys. Commun., vol. 148, no. 2, pp. 160–170, 2002.
- [21] M. Schubert, BNAB 2009: Holography update Television Broadcast, Apr. 20, 2009. [Online]. Available: <http://www.televisionbroadcast.com/article/79134>.
- [22] R. L. D. Valois and K. K. D. Valois, Spatial Vision. Oxford, U.K.: Oxford Science Publications, 1990.

Performance Evaluation of DWDM Systems by Using Various Amplifiers

Amit Kumar Gautam¹, Brahm Raj Singh², Gurjit Kaur³

^{1,2,3}Gautam Buddha University

¹amitkumarg55@gmail.com, ²brahmrajs35@gmail.com,

³gurjeet_kaur@rediffmail.com

Abstract: DWDM system uses optical fiber as a medium for movement of bits between transmitter and receiver. Amplifiers are used in between transmitter and receiver. This paper basically deals with use of different combination of amplifier used in DWDM systems. EDFA amplifier, FRA amplifier and combination of EDFA with FRA can be used in DWDM systems. The various parameters are optical power, optical SNR, BER, Q-factor and noise. Here main focus are on two parameters, first one is optical power and other one is Q-factor.

Keywords: EDFA, FRA, HA, Q Factor

1. INTRODUCTION

In optical fiber communication, Dense wavelength division multiplexing (DWDM) is a technology which uses multiplexing for combining more than one optical carrier signals onto a single optical fiber by using different wavelengths. Dense wavelength division multiplexing (DWDM) basically means the transmission of multiple closely spaced wavelengths by the same fiber.

DWDM systems use normally 100 GHz or even 50GHz, 25 GHz channel spacing for up to 160 channel operation. In this paper three type of amplifiers EDFA, FRA, Hybrid optical amplifier are being analyzed. Optical power is analyzed before amplification and after amplification for EDFA, FRA, Hybrid optical amplifier for 2, 4, 8, 16, 32 number of channels. Q Factor, Eye diagram are being analyzed, effect on performance when number of channels are increased is seen for different amplifiers and result will be compared for sixteen channels.

A. Optical Fiber Amplifiers

1. EDFA

It comes under the category of doped fiber amplifier. This amplifier contains the doping of Erbium ions. The Erbium-doped fiber amplifier (EDFA) is the mostly used fiber amplifier for C-Band (1525 nm -1565 nm) and L-Band (1570nm-1610nm) normally known as third transmission window of silica-based optical fiber.

2. FRA

Fiberramanamplifier (FRA) normally known as Raman Amplifier. This amplifier does the amplification due to the non linear interaction between the signal which has to be amplified and the signal which is obtained from the pump laser.

Hybrid Amplifier

In hybrid amplifier a combination of EDFA and FRA is used for amplification purpose of the transmitted signals. The advantages of both the amplifiers are used to improve the performance of DWDM system. It provides the combination of unique properties of both the amplifiers.

2. SYSTEM DESIGN

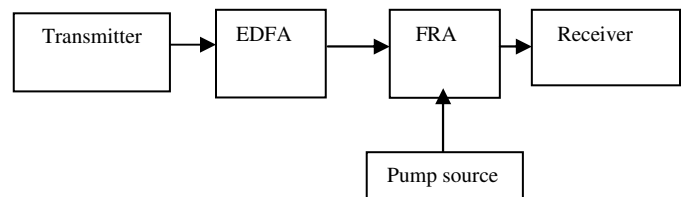


Fig. 1. Hybrid Fiber Amplifier structure diagram

HFA structure diagram shows the Transmitter, Erbium Doped Fiber Amplifier (EDFA), Fiber Raman Amplifier (FRA), Pump Source and Receiver. At Transmitter end a WDM transmitter is used which is able to produce more than one optical signal. After WDM transmitter, WDM Multiplexer will combine the optical signals which are being obtained from WDM transmitter. The multiplexed signal has been transmitted on the optical fiber. Erbium Doped Fiber Amplifier (EDFA) is the amplifier used to amplify the input signal. EDFA are having properties different from FRA. Fiber Raman Amplifier (FRA) is being placed for the amplification of the transmitted optical signal obtained after EDFA. Fiber Raman Amplifier has been excited by a Pump Source. In this paper combination of different amplifiers are being used and evaluation of different parameters of DWDM system is being analyzed. At receiver ideal Demultiplexer is

used to demultiplex the input optical signal. After demultiplexing the signal is converted to electrical signal which is done by photo detector Avalanche Photodiode (APD) and then the electrical signal is being analyzed further.

3. SCHEMATIC LAYOUT

Number of channels which are being considered are 2, 4, 8, 16, 32. But Figure 2 is showing the EDFA system for 16 channels. The frequency of initial channel is 193.1 THz, having frequency spacing of 100 GHz. Modulation type used is NRZ. The length of EDFA is 5 m. At receiver side Demultiplexer is used, optical detector is used to convert the optical signal to electrical signal. In Figure 3 all the parameters are same except optical amplifier, here the amplifier used is Raman amplifier. Figure 4 is the combination of EDFA and Raman Amplifier and can be called as Hybrid optical amplifier system.

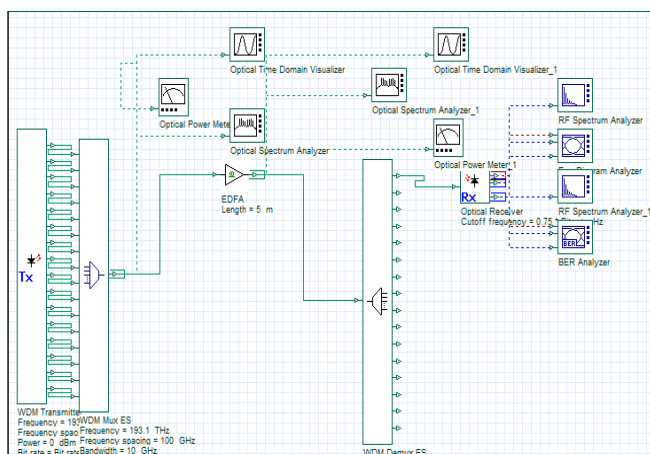


Fig.2. EDFA for 16 channel.

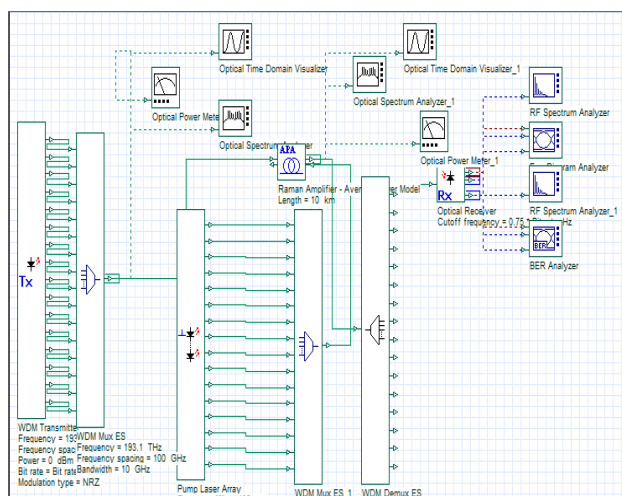


Fig. 3. Raman Amplifier for 16 channel

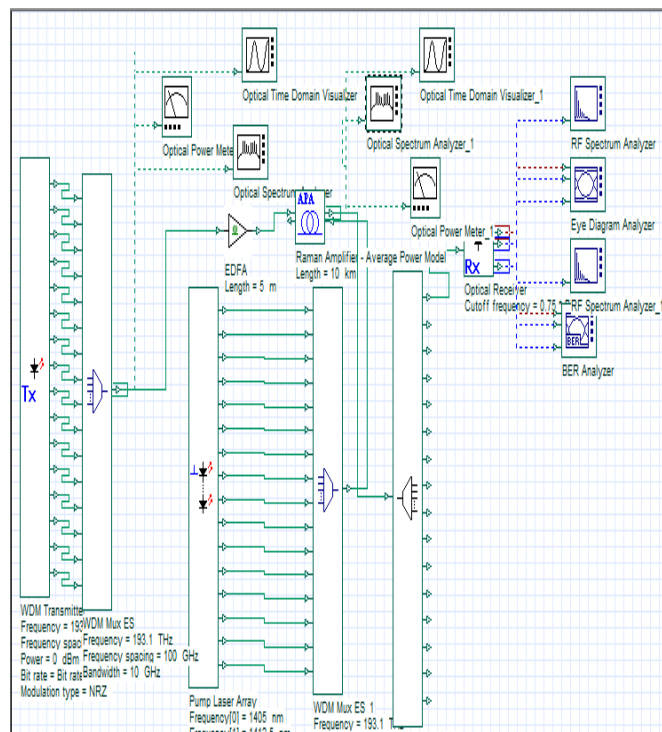


Fig. 4. Hybrid amplifier system for 16 channel

4. RESULT ANALYSIS

Table 1

Number of Channel	Q FACTOR		
	EDFA	FRA	HA
2	2.817	4.815	4.817
4	4.818	4.817	4.819
8	4.841	4.835	4.852
16	4.846	4.844	4.848
32	4.853	4.847	4.859

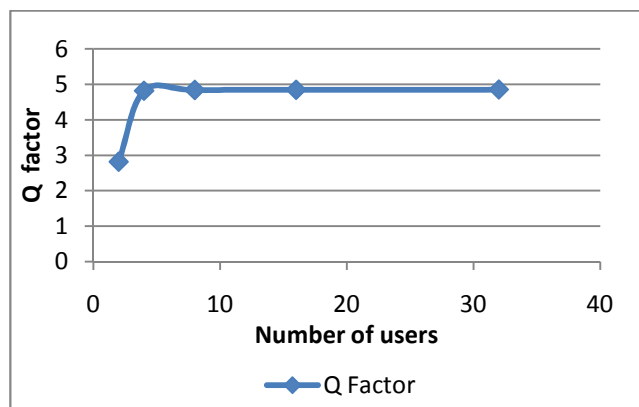


Fig. 5. Number of channels versus Quality factor for EDFA

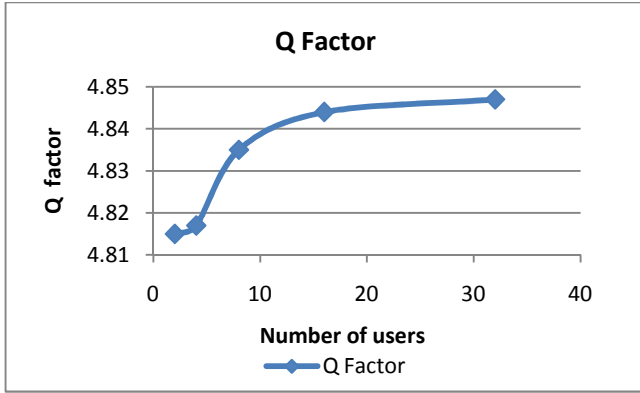


Fig. 6. Number of channels versus Quality factor for FRA

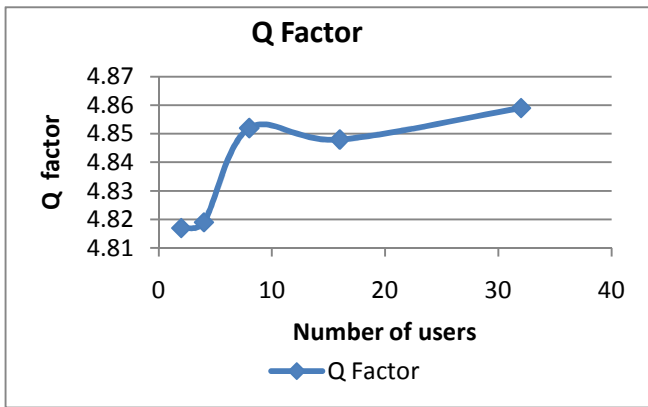


Fig. 7. Number of channels versus Quality factor for HA

Number of channels which are being considered are 2,4,8,16,32. The graph and table has been drawn up to 32 channels. From the Table 1 and the Figure 5, Figure 6, Figure 7 it can be seen that the Q factor of HA is better than FRA and EDFA.

Table 2

Number of Channels	OPTICAL POWER (In dB)					
	BEFORE AMPLIFICATION (In dB)			AFTER AMPLIFICATION(In dB)		
	EDFA	FRA	HA	EDFA	FRA	HA
2	0.117	0.117	0.117	17.542	0.085	17.459
4	3.146	3.146	3.146	17.682	1.145	17.585
8	6.142	6.142	6.142	17.874	4.143	17.735
16	9.148	9.148	9.148	18.185	7.143	17.942
32	12.162	12.162	12.162	18.207	10.567	18.321

HA-Hybrid Amplifier, EDFA-ErbiumDoped Fiber Amplifier, FRA-Fiber Raman Amplifier

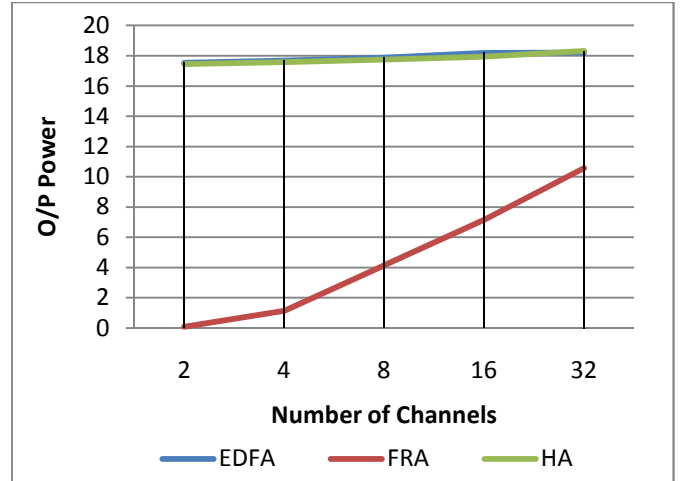
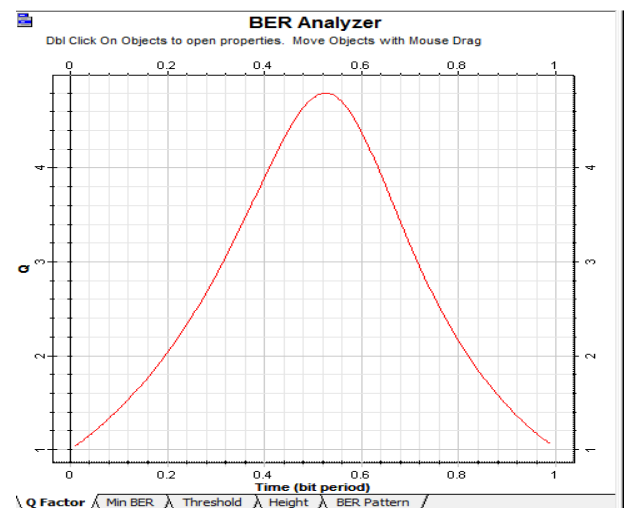


Fig. 6. Output power versus Number of channels

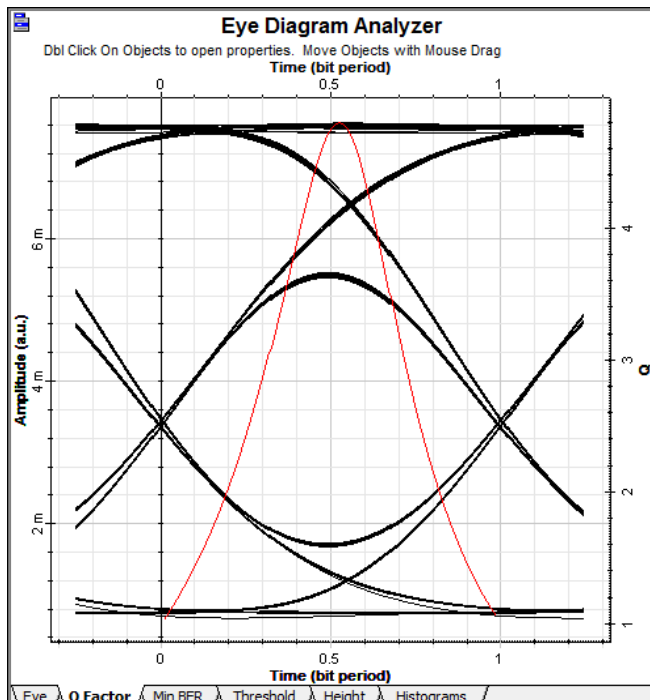
From the Table 2 it can be analysed that the optical power is amplified in each case except FRA. From the Figure 6 it can be deduced that characteristics of EDFA and HA overlap each other because of noise accumulation. But FRA is different because there is no noise accumulation.

5. PARAMETER ANALYSIS DIAGRAM

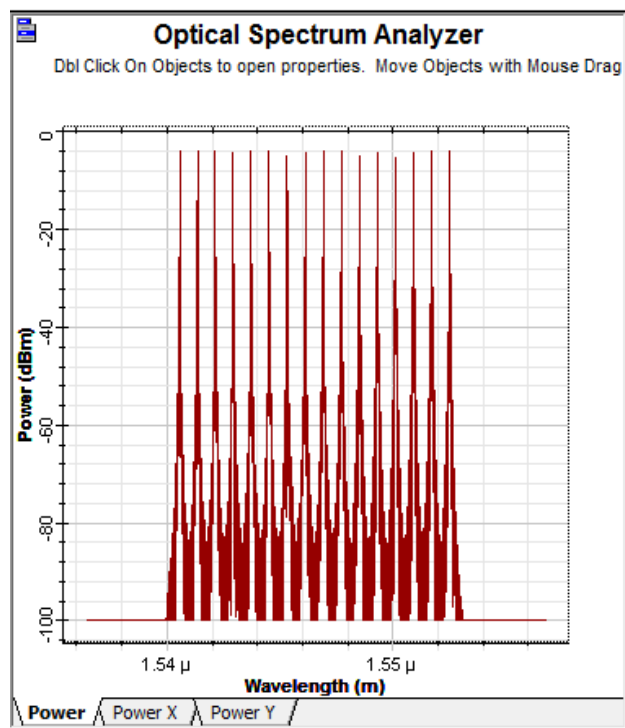
Figure(i) is showing the bit error rate(BER) of HA system for 16 channel which is having maximum value at 0.52 ns. Figure (ii) is showing the eye diagram, eye diagram can be studied as more the opening of eye, means signal received is having good strength and received properly (low distortion is occurred). Figure(iii) shows the multiplexed optical signal, whereas Figure(iv) shows the combination of noise and the optical signal after the amplification.



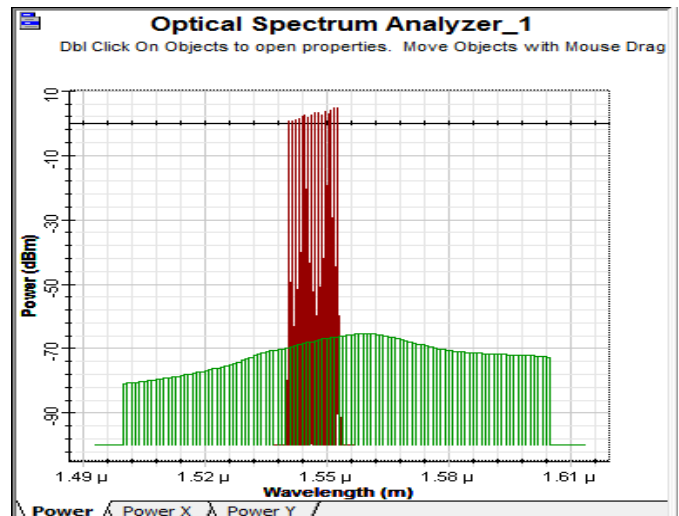
(i)



(ii)



(iii)



(iv)

6. CONCLUSION

The Hybrid Optical amplifiers are the key components for increasing the flexibility and capacity of broadcast optical networks. The performance of optical amplifier and hybrid optical amplifier have been compared. The performance of optical amplifiers was evaluated using the eye patterns, BER measurement, eye opening and Q factor. To achieve better results it is of utmost importance to optimize the optical amplifier. Then the various parameters of hybrid optical amplifier such as Raman pump wavelength, Raman pump power, Raman fiber length, EDFA noise figure and EDFA output power have been optimized.

REFERENCES

- [1] John M. Senior, "Optical Fiber Communication: Principles and Practice", Second Edition, Prentice Hall of India.
- [2] Raman Deep Kaur "Performance Analysis of Hybrid Optical Amplifiers for multichannel WDM systems"
- [3] B. Nagaraju, M.C. Paul, M. Pal, A. Pal, R.K. Varshney, B.P. Pal, S.K. Bhadra, sG. Monnomc and B. Dussardier, 'Design and fabrication of an intrinsically gain flattened Erbium doped fiber amplifier', Journal of Optics Communication, Vol. 282, Pages 2335–2338, 2009.
- [4] Ju Han Lee, You Min Chang, Young Geun Han, Haeyang Chung, Sang HyuckKim, and Sang Bae Lee, 'A Detailed Experimental Study on Single-Pump Raman/EDFA Hybrid Amplifiers Static, Dynamic, and System Performance Comparison', IEEE Journal of Lightwave Technology, Vol. 23, No. 11, Pages 3484-3493, 2005.
- [5] Gerd Keiser, "Optical Fiber Communication: Principles", Fourth Edition, Tata McGraw Hill.

De-De Dodging Algorithm for Scheduling Multiple Instances of Multiple Workflows in Hybrid Cloud

Arun Kumar. B¹, Ravichandran. T², Sundareswari. K³

¹Dept. of CSE, Karpagam University, Coimbatore, Tamil Nadu, arunkumar.oct06@gmail.com

²Principal, Hindustan Institute of Technology, Coimbatore, Tamil nadu

³Dept.of CSE, Karpagam University, Coimbatore, Tamil nadu, sundari88.krish@gmail.com

Abstract: Workflow-based applications usually consist of multiple instances depending on a single workflow, which are jobs with control or data dependencies to provide a well-defined scientific computation task, with each instances acting on its own input data. Due to the raise in convention of many applications currently, there is necessitating for high processing and storage capacity along with the consideration of cost and instance use and also without any deadlocks between those instances. To improve the performance of the entire system a high degree of concurrency is obtained by running multiple instances at the same time. On the other hand, since the amount of storage is limited on most systems, deadlock due to numerous storage requests would-be a problem. In this paper we have proposed a new dependency and deadlock avoidance (De-De algorithm) algorithm along with the consideration of both instance and value. The IVH algorithm that comes to the decision of desiring which resource should be chartered from public providers is now combined with the newly proposed De-De algorithm considering that each instance of both single and multiple workflows should work without any deadlocks. To address this problem, we have combined two new concepts with the traditional problem of deadlock avoidance by proposing a single algorithm that can maximize active (not just allocated) resource utilization and minimize makespan. Our approach is based on the well-known banker's algorithm, but our algorithms make the important distinction between active and passive resources, which is not a part of previous approaches. Through simulation-based studies, we show how our proposed algorithms are better than the classic banker's algorithm.

Keywords: IVH algorithm, De-De algorithm, Scheduling, Multiple workflows, hybrid cloud.

1. INTRODUCTION

Several high-performance computing (HPC) and a set of computations to be completed, such as those already discussed in bioinformatics [1], [2], biomedical informatics [3], cheminformatics [4] and geoinformatics [5], are complicated workflows of single job.

[8] Batch workloads that are typical runs on controlled local area cluster environments. On the other hand organizations that have high workload demands increasingly need ways to share resources across the wide-area, both to lower costs and to increase productivity. One approach to accessing

resources across the wide-area is to simply run a local area batch system across multiple clusters that are spread over the wide-area and to use a distributed file system as a backplane for data access. Alas, this approach is loaded with difficulty, largely due to the way in which I/O is handled. The principal problem in using a traditional distributed file system is in its approach to *control*: many decisions concerning caching, consistency, and fault tolerance are made *implicitly* within the file system. Although these decisions are reasonable for the workloads for which these file systems were designed, they are ill-suited for a wide-area batch computing system.

The workflow is usually organized as a *directed acyclic graph* (DAG), in which the constituent jobs (i.e., nodes) are either control or data dependent (i.e., edges). *Control-flow dependency* specifies that one job has to be completed before other jobs start their process. In contrast, *dataflow dependency* specifies that a job cannot start until all its input data (typically created by previously completed jobs) is available [6]. Control-flow is the more commonly used abstraction to reason about the relationship between different jobs, but we show how dataflow information is more valuable to *effectively* utilize the storage. A workflow-based workload may consist of multiple instances of a workflow. Typically, each instance of the workflow is data-independent of other instances since they compute with different inputs or parameters. [7] Additionally, workflows are collaboratively designed, assembled, validated, and analyzed. Workflows can be shared in the same manner that data collections and compute resources are shared today among communities. The scale of the analysis and thus of the workflows often necessitates that substantial computational and data resources be used to generate the required results. [8] So as a remedy for this, Cloud computing is designed such a way that provides on-demand resources to the users, so as to provide locally available computational power, delivering new computing resources when necessary.

Over the last several years, virtual machines have become a usual deployment object. Virtualization advance enhances flexibility because it abstracts the hardware to the point where software stacks can be deployed and redeployed

without being tied to a specific physical server. Virtualization technology enables a dynamic datacenter where servers provide a pool of resources that are attached as needed, and where the relationship of applications to compute, storage, and network resources changes dynamically in order to meet both workload and business demands.

With application deployment decoupled from server deployment, applications can be deployed and scaled rapidly, without having to first procure physical servers. Virtual machines have become the prevalent abstraction — and unit of deployment — because they are the least-common denominator interface between service providers and developers. Using virtual machines as deployment objects is sufficient for 80 percent of usage, and it helps to satisfy the need to rapidly deploy and scale applications. Virtual appliances, virtual machines that include software that is partially or fully configured to perform a specific task such as a Web or database server, further enhance the ability to create and deploy applications rapidly. The combination of virtual machines and appliances as standard deployment objects is one of the key features of cloud computing.

Cloud Computing vendors combine virtualization (one computer hosting several “virtual” servers), automated provisioning (servers have software installed automatically) and Internet connectivity technologies to provide the service. Consequentially, acquisition costs are low but tenants never own the technology asset and might face challenges if they need to “move” or end the service for any reason. Something that is often overlooked when evaluating Cloud Computing costs is the continued need to provide LAN services that are robust enough to support the Cloud solution. These costs are not always small. For example, if you have 6 or more workstation computers, you will probably need to continue to maintain a server in a domain controller role (to ensure name resolution), at least one switch (to connect all of the computers to each other and the router), one or more networked printers, and the router for the Internet connection.

Basically these are the following types of the Cloud Services: SaaS (Software As A Service) It provides all the functions of a sophisticated traditional application to many customers and often thousands of users, but through a Web browser, not a “locally-installed” application. It eliminates customer worries about application servers, storage, application development and related, common concerns of IT. Highest-profile examples are Yahoo and Google, and VoIP from Vonage and Skype. PaaS (Platform as a Service) Delivers virtualized servers on which customers can run existing applications or develop new ones without having to worry about maintaining the operating systems, server hardware, load balancing or computing capacity. These vendors provide APIs or development platforms to create

and run applications in the cloud – e.g. using the Internet. IaaS (Infrastructure as a Service) delivers utility computing capability, typically as raw virtual servers, on demand that customers configure and manage. IaaS is designed to augment or replace the functions of an entire data center. This saves cost (time and expense) of capital equipment deployment but does not reduce cost of configuration, integration or management and these tasks must be performed remotely. Apart from these we have the following Cloud computing infrastructure models: Public clouds are run by third parties, and applications from different customers are likely to be mixed together on the cloud’s servers, storage systems, and networks. Private clouds are built for the exclusive use of one client, providing the utmost control over data, security, and quality of service. Hybrid clouds combine both public and private cloud models. They can help to provide on-demand, externally provisioned scale. The ability to augment a private cloud with the resources of a public cloud can be used to maintain service levels in the face of rapid workload fluctuations. Sometimes called “surge computing,” a public cloud can be used to perform periodic tasks that can be deployed easily on a public cloud.

2. DE-DE ALGORITHM DESCRIPTION

Workflows $F: \{f_1, f_2, f_3 \dots f_n\}$
 Deadline E
 Resource H
 Predestined Start Value PSV
 Predestined Finish Value PFV
 Public resource pool FB
 Private Resource Pool G
 Rescheduling group N
 Priority Pr
 Pending task PT
 Application Remaining Time ART
 Node set NS
 Time & Cost value TCV
 Job J with the instance i
 Instance of workflows to be scheduled, I_i
 Time taken for completion of a job, $time()$
 Temporary variables W_i and R_i
 Storage request for the job $getWriteSet()$
 Storage allocation of the job $getReadSet()$
 Need of i resources in time t $alloc(i, t) / need(i, t)$
 System safety check $safetycheck()$
 Deadlock Detection Algorithm (DDA),

A. Algorithm

- [1] $F = \text{Set of Workflows} \{ F = \text{Workflows} == \text{set of tasks} \}$
 $TS == \text{single task } T \}$
- [2] function $DDA(I_i, F)$
- [3] $R_i \leftarrow getReadSet()$
- [4] $J \leftarrow J - (|W_i| - |R_i|)$
- [5] $alloc(i, t) \leftarrow alloc(i, t) + (|W_i| - |R_i|)$

```

[6] need (i, t) ← need (i, t) - |Wi|
[7] if (safetycheck (Ii))
[8] J ← J - |Ri|;
[9] alloc (i, t) ← alloc(i, t) + |Ri|;
[10] return true;
[11] goto line 19;
[12] else
[13] J ← J + (|Wi| - |Ri|);
[14] alloc (i, t) ← alloc(i, t) - (|Wi| - |Ri|);
[15] need (i, t) ← need(i, t) + |Wi|;
[16] return false;
[17] goto line 54
[18] End function
[19] Perform initial schedule
[20] Dependency De=0-5
[21] For each W in TW
[22] For each T in TS do
[23] If T < De Do
[24] If (H ∈ G) then
[25] Schedule F in G
[26] While (time(F) > E && iteration =F) do
[27] Select node from NS with ↑Pr
[28] If ni ∉ NS then
[29] Add ni to NS
[30] Iteration=iteration+1
[31] End while
[32] Schedule the H with ↓ PFV
[33] DDA ( Ii, H);
[34] else select next task from TS
[35] else select next workflow from WT
[36] Else
[37] Wi ← getWriteSet ( );
[38] While (|Wi| > G && iteration =F ) do
[39] Request for H in FB
[40] If PFV > ART then
[41] Queue PT to execute
[42] For each W in TW
[43] For each T in TS do
[44] If T < De Do
[45] Select H ∈ FB then
[46] Calculate TCV for new H
[47] If TCV < ( H ∈ G ) then
[48] Add H to FB
[49] else select next task from TS
[50] else select next workflow from WT
[51] Schedule H with ↓ PFV
[52] DDA ( Ii, H);
[53] End while
[54] End else

```

A cloud system receives numerous numbers of requests for a set of resource to complete their jobs. These jobs are termed as workflows. Each of these workflows consists of set of task which in turn is dependent on one another by some means. In this paper the De-De algorithm consider a set of

workflows and detects whether deadlocks occur between them by using the well known banker's algorithm.

The First line of the algorithm initializes the set of workflows that consists of set of tasks T to a variable F. The Function DDA algorithm is defined clearly which includes some of the parameters associated with the instance I_i (i.e., r (t), alloc (i, t) and need (i, t)) are updated accordingly.

In the third line the function DDA is clearly given where R_i is assigned with the allocated resources of the workflows. In the variable G the remaining resource is calculated by subtracting the available resource in the private pool along with the already allocated and requested resources. DDA algorithm first checks if the current available storage is sufficient to satisfy the request of the job (obtained via *getWriteSet()*). If not, the job has to request from the public resource pool. In line seven the safety check algorithm is invoked for verifying whether the system is in safe state or not for each of the workflow. Once verified the line 19 is called if it returns true. In the 19th line initial scheduling is done in which it considers only the Private resource pool and schedule these workflow in the Private resource pool itself based on some attributes like communication cost, priority and time, resource allocation is done. We have assigned a range for dependency for instance: dependency De value is between 0 - 5. The 23rd line checks the range and once if the dependency value is less than the range, the allocation or request to the resource is done else it is not. Next the algorithm checks whether the available resources are enough or not. If it is sufficient enough to finish the job, the workflow is requested in the private cloud itself else it is requested in the public cloud. Once scheduled the workflows in the private cloud, until the deadline is met the task is running inside the private cloud. The iteration is repeated until the deadline E is met, where the algorithm continues by selecting a node N_i from the node set NS with the highest priority. Then the safety check is algorithm is called.

If it returns true then the system is in safe state else system is said to wait and next workflow is considered.

Simultaneously if the resource is not enough in the private pool it is requested in public pool as in line 39. The line 40 in the algorithm verifies whether the Predestined Finish value PFV is greater than ART, then queue the tasks to execute. Again the dependency range is checked for the new and once if the dependency value is less than the range, the allocation or request to the resource is done else it is not. In line 46 evaluate the new TCV for new resource allocation. Once the value of TCV is less than the available resource in the private cloud then only the public cloud is requested. Since the TCV is considered to be less than the old TCV the resource is added to the set NS. Now schedule the resource with the lowest PFV, suppose the TCV value is larger than verify inside the private cloud itself. In line 41 the DDA

algorithm is invoked again for checking safety and if it returns true allocate the resource with the lowest PFV. Finally our algorithm is well furnished to bind between selecting public and private cloud and allocates the requested resources to the particular workflow with the low cost and time and without any occurrence of deadlocks and dependencies between them successfully.

3. RELATED WORK

Deadlock is one of the most discussed problems in the field of operating systems. The theoretical background of this problem as well as its resolution methods have been ingrained and widely deployed since decades ago. As divergent to the traditional batch-oriented workflows, data streaming workflows are continuous and long running in nature, requiring efficient and everlasting transmission of data. The deadlock resolution is particularly vital in these HPC applications because they require high storage cloud be potentially overwhelmed by the incoming data stream if the data arrival rates over take the processing rates but are not properly controlled. Zhang *et al.* [9], [10], has studied this problem and premeditated a suite of repertory strategies to control the start and finish times of the data transfers by setting up upper and lower storage limits. Their storage-aware strategies are based on admission control, a variant of deadlock prevention practice, which is different from ours. As such the recent results in this area are few and far between.

However, in this paper we have provided a case study to show how this problem can be effectively addressed in computational multiple workflows by extending the traditional methods with exploitation of the workflow features. The De-De algorithm attempts to keep the system in safe states, and continues by the use of IVH algorithm, the scheduling process is done by considering both instance and value; effectively provide the selection of choosing between the public and private cloud. Also our paper has included the Banker's algorithm () *a priori* knowledge of the maximum amount of resources needed by each process. Some research efforts focused on refining the banker's algorithm based on some interesting process models, each differing in the amount of information that is assumed to be available. Yu-Kwong Kwok (2004) has made a pair-wise comparison among seven scheduling algorithm under various conditions. But the drawback of this algorithm is that it has a set of several procedures that takes too much time to compile.

A hybrid heuristic scheduling algorithm was implemented on heterogeneous system that comprised of three phases (Sakellariou (2004)). The key idea of the hybrid heuristic is to use a standard list scheduling approach to rank the nodes of the DAG and then use this ranking to assign tasks to groups of tasks that can be subsequently scheduled independently. Rahman, M., (2007). Haluk Topcuoglu has

provided two performance-effective and low complexity task scheduling algorithms namely HEFT and CPOP algorithms for heterogeneous system. Edwin.S.H.Hou has developed a genetic algorithm for multiprocessor scheduling Hou, (1994). The algorithm is based on the precedence relations between the tasks in the task graph. He has compared the genetic algorithm with the list scheduling and optimal schedule using random task graphs and a robot inverse dynamics computational task graphs for various are presented. But this existing algorithm does not provide an optimal solution to the scheme.

4. CASE STUDY

We have made a detailed study using the cloud simulator package. The De-De algorithm is implemented with multiple workflows, the figure 1 shows that the De-De algorithm has improved the storage capacity of the requested resources.

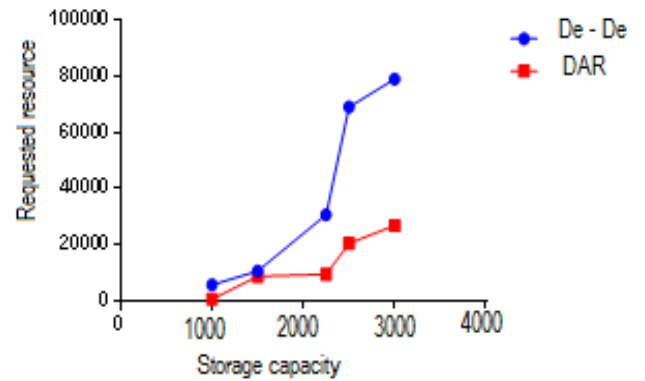


Fig. 1. Performance

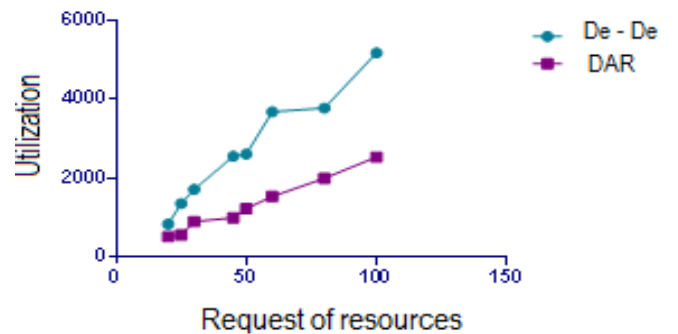


Fig. 2. Utilization factor

The public resources are buffered in a local resource pool, so that it can be used by the other tasks on demand within the leased time. The curve shows that it outperforms the existing DAR algorithm, if the same resources are requested by the multiple tasks possibilities of deadlock occurrence is more, in turn it leads to the increase in access time to specific

resource demanded by the task. Before submitting a workflow to the public cloud, the dependency ratio among the tasks is thoroughly examined using the IVH algorithm and the deadlock between the resources is prevented using DDA algorithm.

The fig 2 shows the utilization factor of multiple resources, the graph shows that multiple resources are effectively used by many tasks without compromising the instance and values. The utilization time of resources on demand is significantly improved using the De-De algorithm, the active storage resources are mapped to the tasks when requested on demand. So by the study made on comparing these algorithms we conclude that our proposed De-De algorithm has improved utilization on the available resource and better find a suitable of selection between public and private clouds.

5. CONCLUSION

Nowadays in many organizations the needs of extra resources are prevailing in a massive range and as a solution to this is the hybrid clouds, which are being used to execute different kinds of applications. Among them, workflows have an important role in processes of many fundamental science fields, such as Physics, Chemistry, Biology, and Computer Science. To speedup science advancements, it is important to provide efficient resource utilization and to execute the service without any deadlocks among them is the major task nowadays. So as a remedy for this in this paper we have designed a De-De algorithm to speed up the execution of multiple workflows obeying a desired execution time and running the system in a safe state by the use of DDA algorithm which is providing us with a better utilization compared to the DAR and IVH approach.

The far-reaching estimation carried out in this work provides sufficient data to support the conclusion that the De-De algorithm can provide efficient scheduling in a hybrid cloud scenario and also maintaining the system in a safe state. Its multicore awareness, along with the cost and time knowledge, can provide makespans as low as the user needs. In general, the proposed algorithm has the ability of reducing the execution costs and time in the public cloud with the increase of the workflow desired execution time. Finally conclude by providing that the DDA method has the potential to achieve better resource utilization because information on the “localized approximate maximum

claims” is used for testing system safety by the use of banker’s algorithm.

REFERENCES

- [1] T. Werner, “Target gene identification from expression array data by promoter analysis,” *Biomolecular Engineering*, vol. 17, pp. 87–94, 2001.
- [2] D. Szafron, P. Lu, R. Greiner, D. Wishart, B. Poulin, R. Eisner, Z. Lu, J. Anvik, C. Macdonell, A. Fyshe, and D. Meeuwis, “Proteome analyst: Custom predictions with explanations in a webbased tool for high-throughput proteome annotations,” *Nucleic Acids Research*, vol. 32, pp. W365–W371, 7 2004, <http://webdocs.cs.ualberta.ca/~bioinfo/PA/>.
- [3] GROMACS, <http://www.gromacs.org>.
- [4] M. Schmidt, K. Baldrige, J. Boatz, S. Elbert, M. Gordon, J. Jensen, S. Koseki, N. Matsunaga, and J. Montgomery, “The general atomic and molecular electronic structure system,” *Journal of Computational Chemistry*, vol. 14, pp. 1347–1363, 1993, <http://www.msg.ameslab.gov/GAMESS/GAMESS.html>.
- [5] B. Ludascher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. Lee, J. Tao, and Y. Zhao, “Scientific workflow management and the kepler system,” *Concurrency and Computation: Practice & Experience, Special Issue on Scientific Workflows*, 2005.
- [6] E. Deelman, D. Gannon, M. Shields, and I. Taylor, “Workflows and e-science: An overview of workflow system features and capabilities,” *Future Gener. Comput. Syst.*, vol. 25, no. 5, pp. 528–540, May 2009.
- [7] A. Ramakrishnan, G. Singh, H. Zhao, E. Deelman, R. Sakellariou, K. Vahi, K. Blackburn, D. Mayers, and M. Samidi, “Scheduling data-intensive workflows onto storage-constrained distributed resources,” in *Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid*, 2007, pp. 401–409.
- [8] J. Bent, D. Thain, A. Arpaci-Dusseau, R. H. Arpaci-Dusseau, and M. Livny, “Explicit control in a batch-aware distributed file system,” in *Proceedings of Networked Systems Design and Implementation (NSDI)*, San Francisco, California, USA, 2004, pp. 365–378.
- [9] W. Zhang, J. Cao, Y. Zhong, L. Liu, and C. Wu, “An integrated resource management and scheduling system for grid data streaming applications,” in *Proceedings of the 2008 9th IEEE/ACM International Conference on Grid Computing*, ser. GRID ’08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 258–265.
- [10] “Block-based concurrent and storage-aware data streaming for grid applications with lots of small files,” in *Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, ser. CCGRID ’09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 538–543.

An Investigation of Cloud Computing

Shubhani¹, Anil Kumar Gankotiya²

^{1,2}Department of CSE, Raj Kumar Goel Institute of Technology for Women, Ghaziabad.

¹shubhi.garg8@gmail.com, ²anilgankotiya@ieee.org

Abstract: Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Internet. It is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand. Cloud computing a relatively recent term builds on decades of research in virtualization, grid computing, utility computing, autonomic computing and more recently networking, web and software services. It implies an architecture which has reduced information technology overhead for the end-user, great flexibility, reduced total cost of ownership, on-demand services and many other things. This paper describes cloud computing, a computing platform for the next generation of the Internet. The paper defines cloud computing, explains the business benefits of cloud computing, and outlines cloud architecture and its major components.

Keywords: Internet, virtualization, autonomic computing, grid computing, utility computing.

1. INTRODUCTION

Cloud computing is the next natural step in the evolution of on-demand information technology services and products. To a large extent, cloud computing will be based on virtualized resources. Cloud computing predecessors have been around for some time now, but the term became popular sometime in October 2007 when IBM and GOOGLE announced a collaboration in that domain. This was followed by IBM's announcement of the "BLUE CLOUD" effort. Since then, everyone is talking about cloud computing. Of course, there also is the inevitable Wikipedia entry.

A. Cloud Computing

Cloud Computing is a paradigm in which information is permanently stored in servers on the internet and cached temporarily on clients that include desktops, entertainment centers, table computers, notebooks, wall computers, hand-held, sensors, monitors, etc. It is a model for delivering information technology services in which resources are retrieved from the internet through web-based tools and applications, rather than a direct connection to a server. Data and software packages are stored in servers. However, cloud computing structure allows access to information as long as an electronic device has access to the web. This type of system allows employees to work remotely.

"Figure1" shows how cloud is acting as an endless pool for storing data and is making the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The cloud-shaped symbol is used as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation.

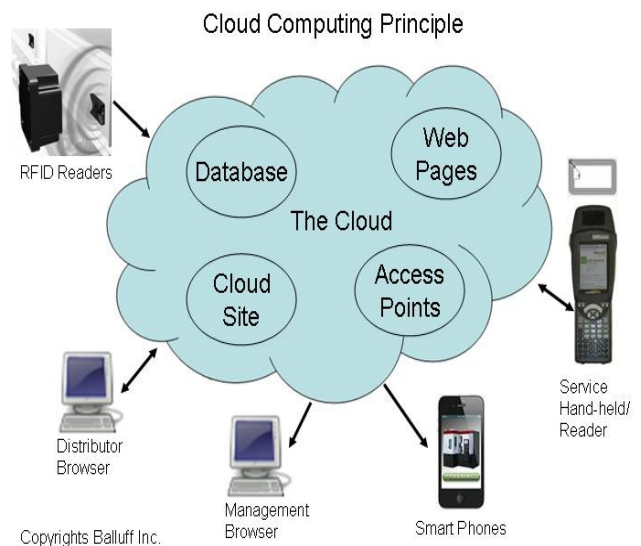


Fig. 1. Cloud Computing

2. CLOUD COMPUTING LAYERS

There are three layers of cloud computing as explained in "Fig.2" which includes SaaS, PaaS, IaaS. These three layers have been described as following:

A. Software as a Service Cloud Model

In this approach, you can rent a service offered by the vendor and then configure the service by using the interface provided by the vendor, without having to know what infrastructure the vendor uses to provide that service. This approach is called *Software as a Service (SaaS)* because you pay to use defined services. For example, Microsoft Exchange Online carries a per-mailbox charge. To configure it, you use a web application supplied by the vendor to request mailboxes, and name and dimension them. You

receive a password for that user and nothing else is necessary—users can access their mailboxes immediately this proposed interface has little in common with the on-premises version of Microsoft Exchange. In a SaaS model, you do not have control over nor are you responsible for the hardware on which the service is installed. Similarly, you have no control over the operating system that runs the service, nor any control over the software apart from what the web user interface exposes to you. In other words, a vendor provides everything required to run the application, shielding you from all the underlying components.

1) Advantages

- Free
- Easy
- Consumer Adoption

2) Disadvantages

- Limited Functionality
- No control and access to underlying technology.

3) Platform as A Service Cloud Model

The next approach is Platform as a Service, or PaaS. In this approach, you rent a platform on which you deploy your applications without configuring the infrastructure and without the limitations of the SaaS approach. You (as a PaaS user) are the developer, building and maintaining the app, so the source code is your responsibility. One interesting perspective is that you can use a PaaS service to build and run SaaS apps. Example-Google App Engine [14] Microsoft Windows Azure [7][12] and Force.com [13]. The key concepts to remember when dealing with PaaS are:

- The platform vendor provides and manages everything, from the network connectivity to the runtime.
- PaaS offerings reduce the developer burden by supporting the platform runtime and related application services.
- Developers can begin creating the business logic for applications almost immediately.

1) Advantages

- Good for developers
- Tightly configured
- More control than application clouds

2) Disadvantages

- Restricted to what is available
- Other Dependencies

C) Infrastructure As A Service Cloud Model

Some vendors provide the infrastructure to build solutions, and you rent the hardware such as servers, load balancers, a firewall, and cables. You then configure these remotely and install your solutions on them. You can scale up by requesting more servers and reconfiguring the load balancer without purchasing more hardware. You can scale down at any time by reconfiguring the infrastructure you rented from the cloud service provider. This vendor approach is called Infrastructure as a Service (IaaS) because a customer can rent the infrastructure without having to forecast and provision for the highest possible demand in advance. In this approach, you are responsible for correctly configuring the rented infrastructure. The obvious benefit of IaaS is that it frees you from the concerns of provisioning many physical or virtual machines. Example-.AmazonEC2 [6][10], GoGrid [9] and FlexiScale [11].

1) Advantages

- Full control of environments
- Control of infrastructure

2) Disadvantages

- Premium Price cost.
- Limited Competition.

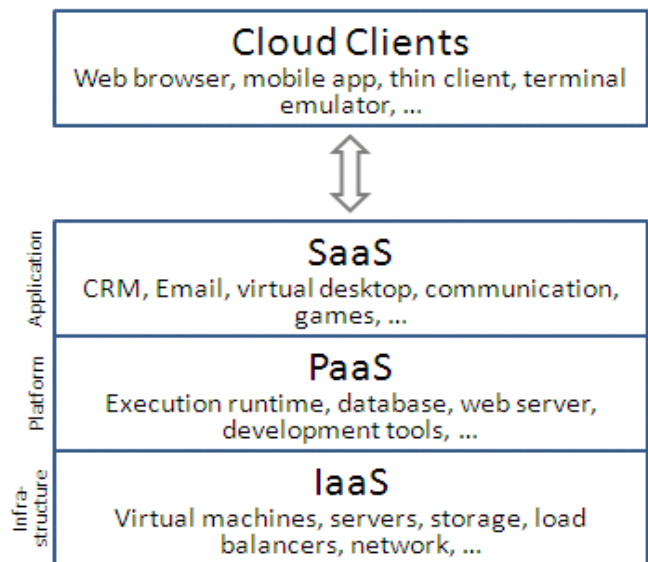


Fig. 2. Cloud Computing Layers

3. CLOUD COMPUTING DELIVERY MODELS

A cloud can be private, public or hybrid. A public cloud sells services to anyone on the Internet. (Currently, Amazon Web Services is the largest public cloud provider.) A private

cloud is a proprietary network or a data center that supplies hosted services to a limited number of people.

A. Public Cloud

A public cloud [3] is established where several organizations have similar requirements and seek to share infrastructure so as to appliance. In addition, it can be economically attractive as the resources (storage, workstations) utilized and shared in the community are already exploited.

This is the cloud computing model where service providers make their computing resources available online for the public. It allows the users to access various important resources on cloud, such as: Software, Applications or Stored data.

A. Advantages

For obvious reasons, public cloud is bound to offer a multitude of benefits for its users, which can be sensed by its ubiquitous demand. Some of the most important ones are mentioned here:

- Efficient storage and computing services.
- Inexpensive, since all the virtual resources whether application, hardware or data are covered by the service provider.
- Allow for easy connectivity to servers and information sharing.
- Assures appropriate use of resources as the users are required to pay only for the services they require.
- Highly reliable and redundant
- Widespread availability irrespective of geographical precincts.
- Sets the business people free from the hassles of buying, managing and maintaining all the virtual resources at their own end, the cloud server does it all.
- Public cloud, in today's advanced workplace, empowers employees and enables them to become productive even when outside the office. The SaaS model ensures that corporations save on IT expenditures while delivering the flexibility of productivity software on the cloud.

2) Disadvantages

- Security is a significant concern in public clouds.

B. Private Cloud

Private cloud is a form of cloud computing where service access is limited or the customer has some control/ownership of the service implementation. It means that either the provider tunnels through that opaque

boundary and limits service access (e.g., to a specific set of people, enterprise or enterprises), or the customer tunnels through that opaque boundary through ownership or control of the implementation (e.g., specifying implementation details, limiting hardware/software sharing). Note that control/ownership is not the same as setting service levels – these are specific to the implementation, and not even visible through the service.

1) Advantages

- They improve average server utilization; allow usage of low cost servers and hardware while providing higher efficiencies; thus reducing the costs that a greater number of servers would otherwise entail.
- High levels of automation reducing operations costs and administrative overheads

2) Disadvantage

IT teams in the organization may have to invest in buying, building and managing the clouds independently.

4. CLOUD COMPUTING ARCHITECTURE

The architecture orients itself around user roles for cloud computing as in “figure-3”. On either end, you have the cloud service creator and cloud service consumer. As its name implies, the cloud service creator role includes any type of cloud service creation tools. These tools include software development environments, virtual, process choreographing solutions, and anything else a developer may use to create services for the cloud.

On the other side of the architecture, the cloud service consumer comes into focus. As you well know, in a cloud environment [8] there are many potential service consumers. The architecture above accounts for in-house IT as well as cloud service integration tools as consumers. There are countless more, but just with these you can begin to appreciate the challenge of effectively enabling the ‘consumer.’ This requires self-service portals, service catalogs, automation capability, federated security, federated connectivity, and more. It is certainly no small task.

Finally, in the middle of the diagram, we have perhaps the most complex role, the cloud service provider. This section builds on top of a shared, usually virtualized infrastructure to address two basic facets for providers: services and service management. From a services perspective [1], we see the trinity of the cloud (IaaS, PaaS, SaaS), with an added wrinkle, Business Process as a Service. As the diagram acknowledges, existing services and partner services will nearly always augment these services, thereby implying the need for tools that provide both functional and non-functional integration capabilities.

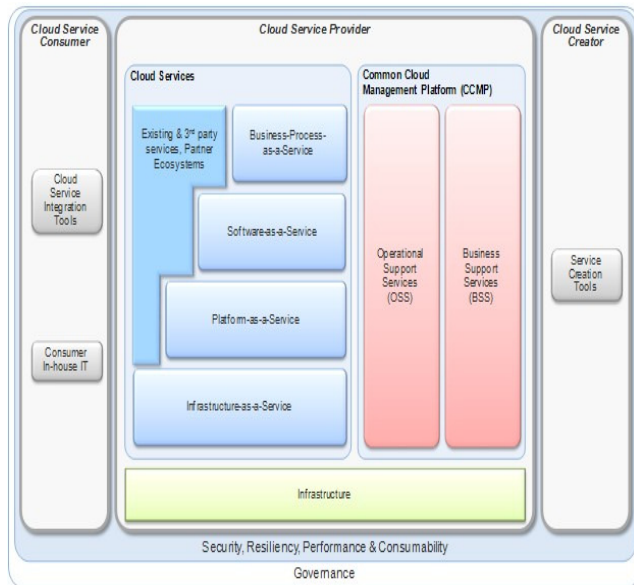


Fig. 3. Cloud Computing Architecture

Opposite the services, we see the common management framework that divides into two major categories: Operational accounts for those capabilities that a provider needs to effectively operate a cloud environment. This includes provisioning, monitoring Support Services (OSS) and Business Support Services (BSS). Naturally, the OSS, license management, service lifecycle management, and a slew of other considerations. BSS outlines the capabilities providers need to support the business requirements of cloud, and this includes pricing, metering, billing, order management, order fulfillment, and more.

5. WHY CLOUD COMPUTING IS IN WORK

Cloud Computing has following features -

A. With context-driven variability, “intelligent assistants” are possible

Because of its expanded [2] computing power and capacity, cloud can store information about user preferences, which can enable product or service customization,” the report states. More than 50% of respondents to the [5] Economist-IBM survey see this as an advantage for addressing fragmented user preferences. A classic example, the authors observe, is Siri, the cloud-based natural-language intelligent assistant on the Apple iPhone 4S. Siri, which enables users to send messages, schedule meetings, place phone calls, find restaurants and more employs artificial intelligence and a growing knowledge base about the user to understand not only what is said but what is meant. “In a nutshell, it leverages the computing capabilities and capacity of cloud to enable individualized, context-relevant customer experiences.”

B. Ecosystem connectivity enables information exchange across business partners.

About a third of survey respondents like how cloud better facilitates external collaboration with partners and customers. Health Hiway, an online health information network that enables the exchange of health information and transactions among healthcare providers, employers, payers, practitioners, third-party administrators and patients in India. “By connecting more than 1,100 hospitals and 10,000 doctors, the company’s software-as-a-service solution facilitates better collaboration and information sharing, helping deliver improved care at a low cost, particularly important in growing markets.”

The examples cited above in the Economist-IBM report are mainly online or tech-savvy companies, but the doors of innovation and disruption are open to all types of businesses from shipping companies to box manufacturers to government agencies. Expect to see a lot of interesting ideas emerge over the coming years, made possible by cloud.

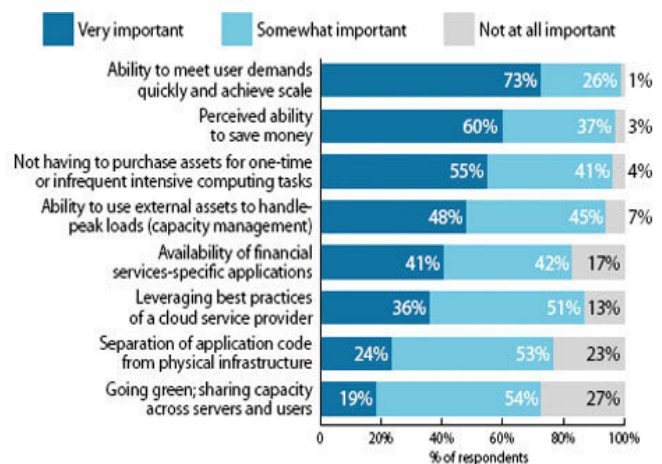


Fig. 4. Importance of Cloud Consideration Factors

The above diagram i.e.”fig-4” explains the importance of the factors which should be considered while switching towards cloud computing. Towards whatever technology you are switching to, all these factors should be well satisfied.

C. Flexibility

Through cloud cost flexibility [4], online marketplace gains access to more powerful analytics online. This is the most obvious and first reason companies are attracted to cloud, and in the Economist-IBM survey, 31% said they like the cloud’s “pay-as-you-go” cost structure. Cloud takes away the need to fund the building of hardware, installing software, or paying dedicated software license fees. This was the appeal for Etsy, an online market place for handmade goods that brings buyers and sellers together and provides recommendations for buyers. “Using cloud -based

capabilities, the company is able to cost-effectively analyze data from the approximately one billion monthly views of its Web site and use the information to create product recommendations,” the report notes. “The cost flexibility afforded through cloud provides Etsy access to tools and computing power that might typically only be affordable for larger retailers.”

D. Scalability

Greater business scalability enables online video retailer to meet spikes in demand: Cloud enables businesses — not just IT operations — to add or provision computing resources just at the time they’re needed. In the Economist-IBM survey, 33% of respondents see this as an advantage of cloud. Netflix, for one, is taking advantage of cloud resources as it meets s up & down demand for its Internet subscription service for movies and TV shows. “Because it streams many movies and shows on demand, the company faces large surges of capacity at peak times,” the report explains. “As Netflix began to outgrow its data center capabilities, the company made a decision to migrate its Website and streaming service from a traditional I data center implementation to a cloud environment. This move allowed the company to grow and expand its customer base without having to build and support a data center footprint to meet its growth requirements.”

E. Adaptability

Greater market adaptability provides online entertainment platform the ability to reach any type of customer device. A third of the executives we surveyed believe cloud can help them adapt to diverse user groups with a diverse assortment of devices. That’s been the experience of Active Video, creator of Cloud TV, a cloud-based platform that unifies all forms of content – Web, television, mobile, social, video-on-demand – onto any video screen, be it set-top boxes, PCs, or mobile devices. “Cloud TV leverages content stored and processed in the network cloud to significantly expand the reach and availability of Web-based user experiences, as well as to allow operators to quickly deploy a consistent user interface across diverse set-top boxes and connected devices,” according to the report. “The Cloud TV approach of placing the intelligence in the network, rather than the device, enables content creators, service providers and consumer electronics manufacturers to create new television experiences for their viewers.”

F. Complexity

Masked complexity enables access to services, no matter how intricate the technology they’re built on: About 20% of respondents cite this as a benefit, demonstrating its still-hidden potential. Because complexity is veiled from the end user, a company can expand its product and service sophistication without also increasing the level of user

knowledge necessary to utilize or maintain the product or service. For example, upgrades and maintenance can be done in the “background” without the end user having to participate. Xerox, through its Cloud Print solution, enables “workers can get their desired content in printed form wherever they might be by using Xerox’s cloud to access printers outside their own organization,” the report says. “While printing from the cloud requires quite a bit of data management – with numerous files to be stored, converted to print-ready format and distributed to printers –the complexity is hidden from users.”

6. TECHNOLOGY RELATED TO CLOUD COMPUTING

Cloud computing typically has characteristics of all these technologies:

A. Grid Computing

Grid Computing involves a network of computers that are utilized together to gain large supercomputing type computing resources. Using this network of computers large and complex computing operations can be performed. In grid computing this network of computers may be present in different locations.

B. Virtualization

Virtualization introduces a layer between Hardware and operating system. During the sixties mainframe started supporting many users using virtual machines. These virtual machines simulated behavior of an operating system for each user. VMware launched a product called VMware Workstation in 1999 that allows multiple operating systems to run on personal computers. The virtualization forms the foundation of cloud technology. Using virtualization, users can access servers or storage without knowing specific server or storage details. The virtualization layer will execute user request for computing resources by accessing appropriate resources.

C. Utility Computing

Utility Computing defines a "pay-per-use" model for using computing services. In utility computing, billing model of computing resources is similar to how utilities like electricity are traditionally billed. When we procure electricity from a vendor, the initial cost required is minimal. Based upon the usage of electricity, electricity companies bills the customer (typically monthly). In utility computing billing is done using a similar protocol. Various billing models are being explored. A few common ones are-

- Billing per user count. As an example if an organization of 100 people uses Google's gmail or Microsoft Live as their internal email system with email residing on

servers in the cloud, Google/Microsoft may bill the organization on per user basis.

- Billing per Gigabyte. If an organization is using Amazon to host their data on the cloud, Amazon may bill the organization on the disk space usage.
- Billing per hour/day. As an example a user may pay for usage of virtual servers by time utilized in hours.

D. Autonomic Computing

Autonomic computing is an initiative started by IBM in 2001. Autonomic means “self-managing” computers. In Autonomic computing, computers can automatically correct themselves without human intervention. As an example consider is a network of computers running a set of programs. When there is a hardware failure on one of the computers on the network, the programs running on that computer are “transferred” to other computers in the network. This is an example of “self-correction” or autonomic computing. The analogy typically used is that of human biological systems. Our biological systems take action in self-correcting mode without our explicit knowledge. In the same way the goal of autonomic computing is for computing infrastructure to self-correct itself in unforeseen situations.

7. PROBLEMS WITH CLOUD COMPUTING

Though from operation and maintenance point-of-view cloud computing is a great cost-effective IT solution for business of any magnitude, but it has two major concerns-technical developments, security and privacy. Since cloud computing is relatively a new technology in comparison to other existing computing solutions, it still has lots of scope of becoming a mature system as a reliable and cost-effective computing technology.

Since due to outsourcing all the important data resides in a third party premise, there is always a concern about the trustworthiness of the cloud service providers. Any security and privacy violation can be fatal- keeping this in mind many business owners are still to be convinced about the security and privacy issues of cloud computing.

More sensitive data are banking and Governmental data. Just think about a classified document of any Govt. agency getting leaked or user’s credit card information falling into the hands of cyber criminals. The only REMEDY is the cloud privacy issues should be more and more organized and strict rules and governance for cloud operation should be implemented, so that, the more and more business will feel safe to opt for cloud computing.

Despite some serious privacy related drawbacks, cloud computing is a lucrative choice to improve productivity in

any business environment, where IT is in high demand. To raise the security and privacy of cloud service providers, there need to be more co-operations between world governments so as we can develop a unified global rules and guidance for running a safe cloud computing service.

8. CONCLUSION

Cloud computing has recently emerged as a compelling paradigm for managing and delivering services over the Internet. The rise of cloud computing is rapidly changing the landscape of Information Technology, and ultimately turning the long-held promises of Utility Computing into a reality. However, despite the significant benefits offered by cloud computing, the current technologies are not matured enough to realize its full potential. Many key challenges in this domain, including automatic resources provisioning, power management and security management, are only starting to receive attention from the research community. Therefore we believe there is still tremendous opportunity for researchers to make groundbreaking contributions in this field and bring significant impact to their development in the industry. In this paper, we have discussed cloud computing, covering its essential concepts, architectural designs, prominent characteristics, key technologies as well as challenges. We hope our work will provide a better understanding of cloud computing.

REFERENCES

- [1] Lewis, Grace. Cloud Computing: Finding the Silver Lining, Not the Silver Bullet.
- [2] Lewis, Grace. Basics about Cloud Computing.
- [3] Jansen, WAYNE & Grance, Timothy. Guidelines on Security and Privacy in Public Cloud Computing. National Institute of Standard and Technology, 2011.
- [4] Strowd, Harrisor & Lewis, Grace. T-check in System-of-Systems Technologies: Cloud Computing (cmu/sei-2010-tn-009). Software Engineering Institute, Camegie Mellon University, 2010.
- [5] IBM homepage. IBM and Google announce university initiative to address internal- scale computing challenges.
- [6] Amazon Web Services. Amazon Elastic Compute Cloud homepage.
- [7] Chappel D. Microsoft Azure homepage. Introducing the Azure Services Platform.
- [8] Hand, Eric. “Head in the Clouds”. Nature .25:449(2007- oct).
- [9] Cloud Hosting, Cloud Computing and Hybrid Infrastrure from GoGrid <http://www.gogrid.com>.
- [10] Amazon Elastic Computing Cloud, aws.amazon.com/ec2.
- [11] Flexiscale Cloud Computing and Hosting, www.flexiscale.com
- [12] Windows Azure, www.microsoft.com/azure.
- [13] Salesforce CRM, <http://www.salesforce.com/platform>.
- [14] Google App Engine, <http://code.google.com/appengine>.

Analysis of MEMS Application

Megha Goyal¹, Dolly Gupta²

Dept. of Electronics & Communication, Dronacharya College of Engineering, Gurgaon, India

¹meghagoyal2010@gmail.com, ²dolly.professional@gmail.com

Abstract: MEMS or Micro-Electro-Mechanical Systems are chips that are made in semiconductor fabrication plant combining electronic functions and mechanical actions. MEMS accelerometers are used to sense the acceleration experienced by a system. The development of MEMS accelerometers has been driven by the demand of the automobile industry for an inexpensive accelerometer as an airbag sensor, computer and audio-video technology. MEMS accelerometers are much smaller, lighter, and more reliable and are produced for a fraction of the cost of the conventional bulky accelerometers. In this paper, we will describe the rapidly emerging field of MEMS accelerometer and discuss its present and future applications.

Keywords: MEMS, Accelerometer, Micromachining, Automobile, Airbags

1. INTRODUCTION

An accelerometer is an electromechanical device that can measure the force of acceleration, whether caused by gravity or by movement. These forces may be static, like the constant force of gravity pulling at our feet, or they could be dynamic - caused by moving or vibrating the accelerometer. An accelerometer can therefore measure the speed of movement of an object it is attached to. There are many types of accelerometers developed and reported in the literature. As devices are getting smaller and smaller day by day, something smaller had to be tried in the field of electronics that could increase applicability. Hence were developed MEMS (Micro Electro-Mechanical Systems) accelerometers. The first micro machined accelerometer was designed in 1979 at Stanford University, but it took over 15 years before such devices became accepted mainstream products for large volume applications [1]. In the 1990s, MEMS accelerometers revolutionised the automotive-airbag system industry. Since then they have enabled unique features and applications ranging from hard-disk protection on laptops to game controllers. Micro machined accelerometers are a highly enabling technology with a huge commercial potential. They provide lower power, compact and robust sensing. Multiple sensors are often combined to provide multi-axis sensing and more accurate data [2].

2. WHAT IS MEMS?

The term MEMS refers to the micro-scale 'components' or micro-scale 'devices' within a system, not the entire system

itself. In order to create a completely functioning 'System' that makes use of MEMS, the system will require various other sub-systems, such as: power, microelectronics, communication and software, as illustrated in fig. 1 below. MEMS are fabricated with a unique set of technologies collectively referred to as 'micro fabrication' or 'micromachining'. These methods are quite different from macro-scale techniques. Due to their small size, standard machine tools cannot be used to machine MEMS features. Micromachining technology is closely related to IC (integrated circuit) fabrication, with some notable differences. There are two main areas of micromachining:

- Surface Micromachining, which is based on the successive deposition and etching of thin films of material such as silicon nitride, polysilicon, silicon oxide and gold.
- Bulk Micromachining, which is based on the etching and bonding of thick sheets of material such as silicon oxides and crystalline silicon.

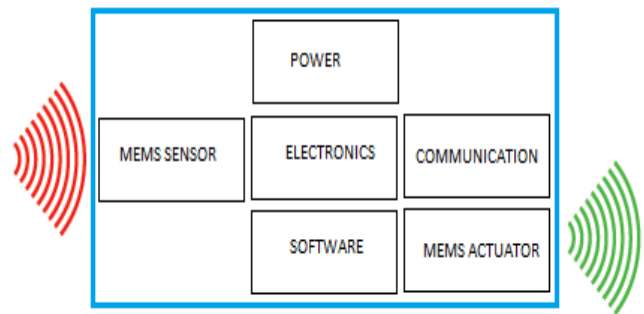


Fig. 1. MEMS Block Diagram

3. MEMS ACCELEROMETER

The early commercial MEMS accelerometers were of the piezoresistive type realized by silicon bulk micromachining. MEMS accelerometers had their first major commercial success in airbag crash sensing in automobiles [3]. The simplest MEMS accelerometer is an inertial mass suspended by springs as depicted in fig.2 below. The scale shows the acceleration along the sensitivity axis. Unit g is acceleration subject to all bodies at the surface of the earth due to gravity, equal to 9.8 meter/second². Deflection of the mass is converted into an electrical signal.

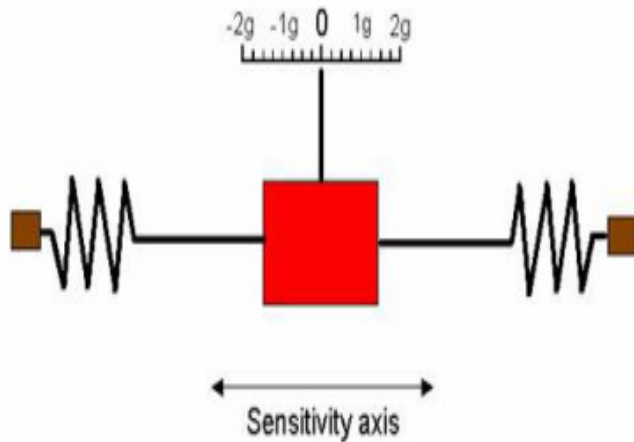


Fig. 2. Schematic Structure of an Accelerometer

MEMS accelerometers use nanotechnology in order to enhance the natural abilities common between all accelerators; hence, these devices are extremely fine-tuned and accurate. Accelerometer usually contains some type of spring force in order to balance the external pressure and displace its mass, thus leading to the motion that produces acceleration. MEMS accelerometers are actually the simplest type, since they consist of little more than a seismic mass, also known as a proof mass, as well as a cantilever beam. Fig.3 shows piezoresistive & capacitive based MEMS accelerometer design. External force is applied which shifts the position of the proof mass from a neutral position to an active position; typically the amount of this deflection is measured by analog or digital readouts. The variations can be charted by using a set of beams that are fixed in place contrasting with a set of beams that have been attached to the surface of the proof mass somehow. Such a simple system makes the accelerometer not only reliable but also relatively inexpensive to manufacture. Top-notch MEMS accelerators are built with quantum tunneling in order to achieve the highest sensitivity possible. [4] These accelerators are accurate enough that they can be measured optically. Most accelerometers function on one axis, but two-axis and three-axis models have also been invented. The three-axis model is naturally more expensive but also far more accurate. Usually higher the device can measure, more the accuracy suffers. Hence, there has to be trade off between the amount of measurement and accuracy involved.

There are two primary sensing techniques employed in MEMS accelerometers - piezoresistive and capacitive. Piezoresistive devices exhibit a change in resistance with respect to applied acceleration. These units tend to be more rugged and are used for accelerometers that achieve higher amplitudes and higher frequency response. On the other hand, capacitive units offer higher sensitivities and are utilized for low amplitude, low frequency devices.

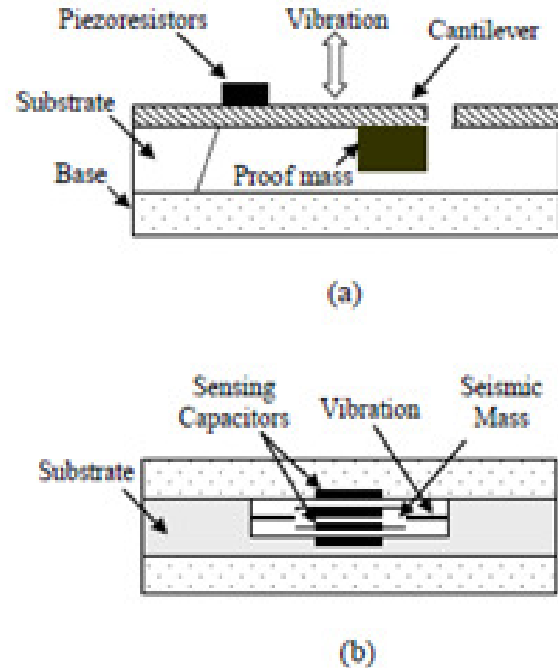


Fig. 3. A typical MEMS accelerometer construction; (a) piezoresistive using cantilever design, (b) capacitance based

4. DESIGN PARAMETERS FOR MEMS ACCELEROMETER

The most common MEMS accelerometer design parameters are resolution, sensitivity (scale-factor), bandwidth, nonlinearity, bias drift, and scale factor asymmetry. The design parameters of MEMS accelerometers are shortly explained below.

Resolution is defined as the noise floor of the accelerometer system. The noise sources in accelerometer system come from both mechanical and electrical parts of the system [5]. The unit of the resolution is $g/\sqrt{\text{Hz}}$ or g in a fixed bandwidth where g is the earth gravity. Sensitivity or scale-factor is the change in the response of the system to $1g$ acceleration. The unit of sensitivity can be F/g if the mechanical part sensitivity is considered only or V/g if mechanical part and readout circuitry is considered together. Bandwidth is the length of the frequency range that input signal frequency can vary. The bandwidth values are directly related to resonance frequency of the accelerometer. Nonlinearity of the accelerometer is defined as the deviation of the accelerometer response from the best fit curve for different magnitudes of acceleration signal in its working range. Bias drift is defined as the maximum deviation of the accelerometer system output with time for a fixed input acceleration signal. Scale factor asymmetry is another parameter that is directly related to fabrication. The definition of the scale factor asymmetry is the percentage difference of scale factor best fit values for negative and

positive acceleration inputs. Other design parameters, which also affect the performance of the accelerometers, are temperature dependence of each other parameter, axis misalignment, cross-axis nonlinearities, and high order nonlinearity constants. Considering all the design parameters an accelerometer model can be constructed as [6]

$$E_0 = K_1(K_0 + a_i + K_2a_i^2 + K_3a_i^3 + J_0a_p + J_p a_0 + K_{ip}a_i a_p + K_{io}a_i a_o)$$

where

E_0 = accelerometer output

a_i, a_p, a_o = accelerometer inputs for each axis

K_0 = bias

K_1 = scale factor

K_2 = 2nd order nonlinearity

K_3 = 3rd order nonlinearity

J_0, J_p = misalignment

K_{ip}, K_{io} = cross axis nonlinearities

In this study scale factor, bias, total nonlinearity, and resolution are considered as design parameters. The misalignment in the axis is packaging related problem and is not considered as a design parameter.

5. APPLICATION AREAS

MEMS Accelerometers are finding applications in numerous fields ranging from consumer devices to military & industry. They are being used as tilt sensors for tagging the orientation of the handheld smart device to sealing the hard disk drive when the device accidentally falls down. In automobile crashes, the accelerometer would detect the rapid negative acceleration and deploys the airbag at just the right time. Table 1 below lists the present application areas of MEMS accelerometers widely in use. [7][8][9]

6. CURRENT USES OF MEMS ACCELEROMETER

Application Areas	Product	Features & Functions
Personal Electronic Devices	Media Players, Gaming devices, Smartphone's (iphone, N95), Contactless Game Controllers, Mouse	Step Counters, User interface control, Switching between portrait and landscape modes.
	Camcorders	Image stabilization
	Digital Still cameras	Anti blur capturing
	Laptops	Hard Disk Drive (HDD) protection
Automotive Industry	Aircrafts, Cars etc.	Airbags for crash sensing detection & protection

Application Areas	Product	Features & Functions
Military & Aerospace Systems	Smart Weapon Systems	Direct and indirect fire, Aviation Launched and Ship Launched Missiles, Rockets, Projectiles
Medical	Pedometer	To calculate speed and distance in shoes.

One of the crucial uses for MEMS accelerometers, in particular, has been airbag deployment systems; they literally save lives because they are able to judge when two cars have struck each other and even ascertain the severity of the collision, adjusting airbag size and rate of deployment accordingly. Another crucial application is protection of hard disk drives in digital equipments.

An airbag is a vehicle safety device. It is an occupant restraint system consisting of a flexible fabric envelope or cushion designed to inflate rapidly during an automobile collision. Its purpose is to cushion occupants during a crash and provide protection to their bodies when they strike interior objects such as the steering wheel or a window.

Modern vehicles may contain multiple airbag modules in various sides and frontal locations of the passenger seating positions, and sensors may deploy one or more airbags in an impact zone at variable rates based on the type, angle and severity of impact. The airbag is designed to inflate in moderate to severe frontal crashes. A central "Airbag Control Unit" monitors a number of related sensors within the vehicle including accelerometers, impact sensors, side (door) pressure sensors, wheel speed sensors, gyroscopes, brake pressure sensor and seat occupancy sensors [10]. The airbag sensor is a MEMS accelerometer, which is a small integrated circuit with integrated micro mechanical elements. The microscopic mechanical element moves in response to rapid deceleration, and this motion causes a change in capacitance, which is detected by the electronics on the chip that then sends a signal to fire the airbag. Fig.3 & 4 shows a schematic of the airbag and how it inflates in an event of collision.

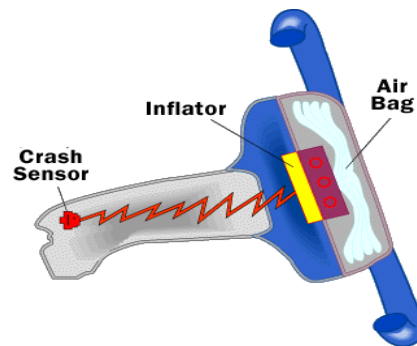


Fig. 3. Airbag

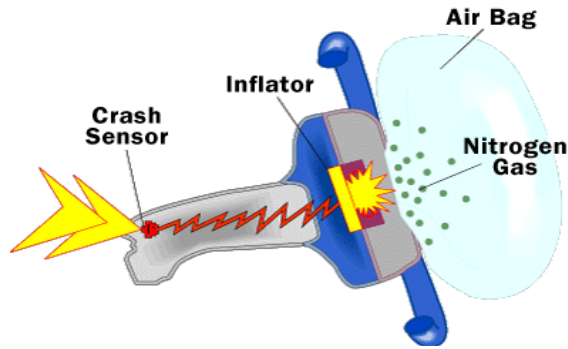


Fig. 4. Inflated Airbag

Hard disk drives (HDDs) are becoming more widely used than ever before, due to the explosive growth in the introduction of portable equipment such as laptop computers, portable media players (PMPs), and handsets. As more and more devices incorporate HDDs, the need has become more pressing to protect them from shocks produced by severe impacts when a product that contains one is dropped accidentally and especially, when this impact occurs with HDD in a read or write mode at the instant fall begins. To increase the ability of HDDs to survive such events, their impact resistance must be enhanced. There are two approaches to establishing the necessary impact resistance, active and passive. Passive approaches have been in use for a long time, they simply cushion the device with impact-absorbing materials usually rubber or gels. Among active approaches, there are two alternatives for protecting HDDs. One is to increase cache memory capacity so that the HDD is in a read or write mode less often. The second approach is to employ accelerometers which measure axial acceleration to detect a drop and then generate a signal that causes an HDD head to be recalled to a safe zone. If this can occur before the product hits the floor or other stationary surface, a collision between head and platter will be prevented. This approach was first used commercially in a notebook PC in 2003 called the APS (Active Protection System). A parked head has much less chance of damaging data than if the head is over sectors containing data when impact occurs.

7. FUTURE SCOPE

Growth till date in the field of MEMS Accelerometers has come from a combination of technology displacement, as exemplified by automotive pressure sensors and airbag accelerometers and new products, such as miniaturized guidance systems for military applications and wireless tire pressure sensors. Much of the growth in this business is expected to come from products that are in the early stages of development or yet to be invented. Some of these devices include disposable chips for performing assays on blood and tissue samples, which are now performed in hospital laboratories, integrated optical switching and processing chips, and various RF communication and remote sensing

products. The development of micro (less than 100kg) and nano (about 10kg) satellites is bringing the mass and volume advantage of MEMS to good use [8]. Other fields that are regularly employing and developing accelerometers include navigation, transportation, consumer electronics, and structural integrity which feed into construction, architecture, and other building-related trades. Accelerometers are increasingly gaining popularity with marine biologist and animal scientists. By measuring the Overall Dynamic Body Acceleration, animal's behavioural patterns can be tracked and scientists can discover how much energy an animal uses in the wild and how quickly they expend that energy [2]. The key to enabling the projected 25-fold growth in MEMS products is development of appropriate technologies for integrating multiple devices with electronics on a single chip [11].

8. CONCLUSION

MEMS devices are used in virtually all areas of industrial activity, health care, consumer products, construction, military and space hardware. MEMS industry in many aspects is still a young industry. Size of MEMS is getting smaller, frequency response and sense ranges are getting wider. MEMS are getting more and more reliable and their sensitivity better every day. We can be sure that the future for MEMS is bright.

REFERENCES

- [1] I. Lee, G. H. Yoon, J. Park, S. Seok, K. Chun and K. Lee, "Development and analysis of the vertical capacitive accelerometer", *Sensors and Actuators A* 119, pp. 8-18, 2005
- [2] S. Beeby, G. Ensell, M. Kraft. and N. White, "MEMS mechanical sensors" (Artech house inc., USA, 2004)
- [3] L.M. Roylance and J. B. Angell, "A Batch-Fabricated silicon accelerometer", *IEEE Trans. Elec. Dev.*, ED-26, 1911 (1979)
- [4] P. Walter, "The History of the Accelerometer", *Sound and Vibration Magazine*, page no. 84, (January 2007)
- [5] J.R. Vig and Y. Kim, "Noise in microelectromechanical system resonators", *IEEE Trans. on Ultrasonic, Ferroelectrics, and Frequency Control*, Vol.46, No.6, pp.1558-1565, Nov. 1999
- [6] IEEE standard specification format guide and test procedure for linear, Single axis, Pendulous, Analog torque balance accelerometers, 1972
- [7] <http://en.wikipedia.org/wiki/Accelerometer> (14.2.2008)
- [8] F. Chollet, H. Liu, "short introduction to MEMS" (18.2.2008)
- [9] http://www.sensormag.com/articles/0399/0399_44/main.shtml (14.2.2008)
- [10] M.Perlmutter, L.Robin, "High performance, low cost inertial MEMS: A market in motion", *Position Location and Navigation Symposium (PLANS)*, pp. 225-229, IEEE 2012.
- [11] M.Trifunovic, A.M.Vadiraj, van Driel, W.D. "MEMS accelerometers and their bio-applications" 13th international conference on thermal, mechanical and multi-physics simulation and experiments in Microelectronics and Microsystems (EuroSimE), pp.1/7 - 7/7, 2012

4G: A New Regime in the Mobile Communications Generations

Rekha Kashyap¹, Ankush Kapoor²

¹B.Tech 3rd Year ECE Department, Jawaharlal Nehru Government Engineering College, Sundernagar Distt. Mandi H.P, Pin-175018, bhandarirekha28@gmail.com

²Assistant Professor ECE Department, Jawaharlal Nehru Government Engineering College, Sundernagar Distt. Mandi H.P, Pin-175018, ankush8818@yahoo.com

Abstract: Fourth Generation an upcoming standard being designed to allow communication between wireless devices across many different wireless standards. This paper presents an analysis of fourth generation wireless features. It discusses what fourth generation wireless is able to accomplish and compares it to the previous generations of wireless technology. This paper will also give a brief overview of benefits and setbacks of fourth generation. The paper ends with a look into the challenges for fourth generation wireless and what lies ahead

Keywords: Fourth generation, generations, multi input multi output, comparison, challenges.

1. INTRODUCTION

Fourth Generation (4G) is the next technological strategy in the field of wireless communication. It is expected to upgrade the existing communication networks. It will also provide a secure IP based solution where facilities such as voice, data and streamed multimedia will be provided to the users on a “Anytime, Anywhere” basis and at much higher data rate compared to previous generations. Cellular providers have the opportunity to offer data access to a wide variety of devices. The cellular network would become a data network on which cellular phones could operate-as well as any other data device. Sending data over the cell phone network is a lucrative business. In the information age, access to data is the “killer app” that drives the market. The most telling example is growth of the internet over the last ten years. Wireless provides a unique twist to this product: mobility.

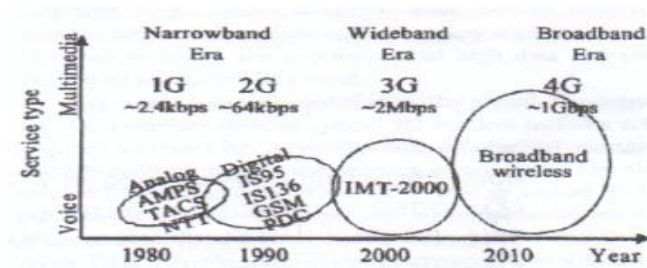


Fig. 1. Evolution of Cellular System [6]

2. GENERATIONS

Today “mobile has become the part and parcel of our life”. And as we being humans move from one generation to the next generation similarly the mobile communication has its different generations. Advancement came with different generations. Generations are as follows: first generation, second generation, third generation, fourth generation. Now let us take a dive into the world of mobile and different technology evolutions pertaining to it.

(2a). First Generation (1G):

Advanced Mobile Phone System (AMPS). It was first launched in UNITED STATES (US). It is an analog system based on FDMA (Frequency Division Multiple Access) with 30kHz FM modulated voice channels. Spectrum allocated by FCC (Federal Communication Commission) is 50MHz. [1] Today, it is the most used analog system and second largest worldwide. Total Access Communication System (TACS) first used in the UK in 1985.

(2b). Second Generation (2G):

Global System for Mobile Communications (GSM) was the first commercially operated digital cellular system. It was developed in 1980s [2]. The European Telecommunication Standards Institute was responsible for the standardization of GSM. GSM uses TDMA (Time Division Multiple Access) and slow frequency hopping with frequency shift keying for voice modulation. It is the dominant cellular standard today, with over (45%) of the world's subscriber at April 1999. There are two standards in the 900 Mhz cellular frequency band as follows: IS-136 is the digital enhancement of the analog AMPS technology. It uses a combination of TDMA and FDMA (Frequency Division Multiple Access) and phase shift keyed modulation. IS-95 uses direct sequence CDMA (Code Division Multiple Access) with phase shift keyed modulation and coding. It increases the capacity by using the entire radio band with each using a unique code.

(2c). Two Five generation (2.5G)

The second generation (2G) digital cellular standards have been enhanced to support high rate packet data service. GSM

provides a data rate up to 140kbps by aggregating all time slots together for a single user and this enhancement is called as General Packet Radio Switch(GPRS).A more Fundamental Enhancement, EDGE- Enhanced Data Rate For GSM Evolution. Provides data rate up to- 384kbps using high level modulation format and coding.IS-136 uses GPRS and EDGE enhancement to support data rate up to 384kbps. IS-95-supports data rate upto-115kbps

(2d). Third Generation (3G)

The concept of third generation is based on wideband CDMA standard development under auspices of ITU (International Telecommunication Unit). International Mobile Telecommunications2000 (IMT-2000) is expected to provide different data rates depending on mobility and location. For pedestrian use-384 kbps, vehicular use-144kbps, indoor office use-2mbps.The spectrum between 400MHz and 3 GHz is technically suitable for the third generation.IMT-2000 also provided the following key advantages: 1.flexibility 2.Affordability 3.Compatibility with existing system 4.Modular design. Its key vision is to provide seamless global roaming, enabling users to move across borders while using the same number and handset.

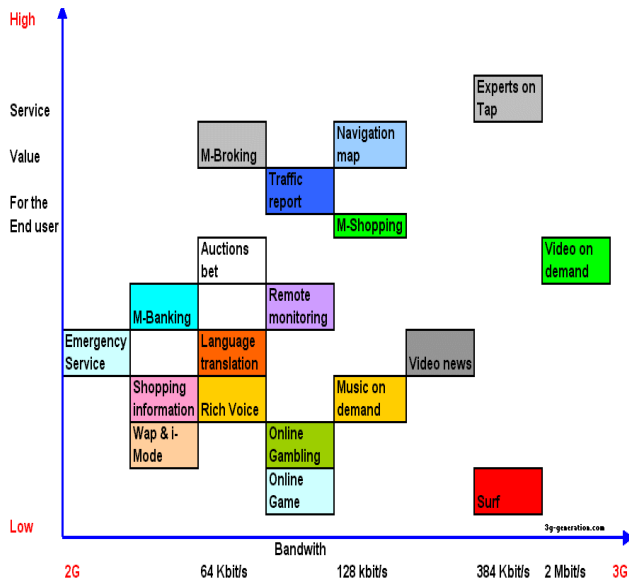


Fig. 2. 3G Applications, Services and Market [6]

(2e). Fourth Generation (4G)

4G cellular systems have targeted peak data rate up to approximately 100Mbit/s for high mobility such as mobile and data rate up to 1Gbit/s for low mobility such as local wireless access, according to the ITU requirement. Bandwidth up to at least 40MHz should be provided. A 4G system is expected to provide a comprehensive and secure all-IP based solution where facilities such as IP telephony, ultra broadband internet access, gaming services and HDTV

streamed multimedia may be provided to users. The infrastructure for 4G will be only packet based (all-IP).

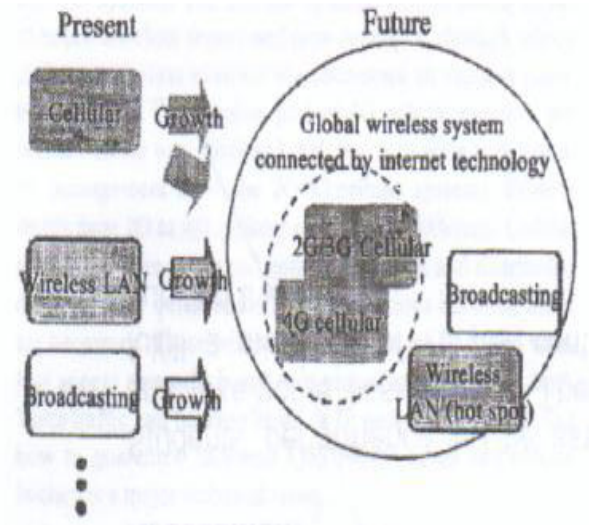
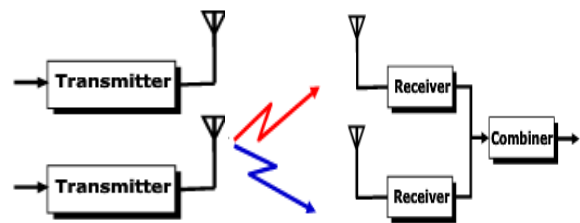


Fig. 3. Global 4G Wireless System[6]

Multi Input Multi Output (Mimo)

When well developed adaptive antenna array were introduced into mobile communication, smart antennas came into existence. It was around early 1990s.It provides better signal and frequency usage for wireless communication. One way of categorizing smart antennas is on the basis of number of inputs and outputs that is used for the device. The classification is as follows: SIMO (single input, multiple output), MISO (multiple input, single output), and MIMO (multiple input, multiple output).



MIMO Circuit

Fig. 3. MIMO Circuit

3. FEATURES OF 4G WIRELESS SYSTEMS

1. High network capacity i.e.the number of simultaneous users per cell will be more.
2. Handoff across different networks will be smooth.
3. Seamless connectivity
4. Global roaming across multiple networks will be possible.

5. A nominal data rate of 100Mbit/s when the customer moves physically at high speed relative to the station, and 1Gbit/s when the customer and station are in relatively fixed positions as defined by the ITU-R
6. A data rate of 100 Mbit/s between any two points of the world
7. An all IP, packet switched network.
8. Interoperability with the existing wireless standard.
9. Quality of service high for next generation multimedia support: high speed data, HDTV video, mobile TV, real time audio and many other.

4. COMPARING PARAMETERS OF 4G WITH CURRENT TECHNOLOGY

Comparison of 4G with the current scenario will help us better understand what actually the implementation of 4G lead to and what all it will take to implement the fourth generation .

	3G	4G
Major requirement driving architecture	Dominantly voice driven-data was always additional	Converged data and voice over IP
Network architecture	Wide area cell based	Hybrid Integration of Wireless LAN and wide area
speeds	384 kbps to 2mbps	20 to 100 mbps in mobile mode
Frequency band	depends on country or continent(1800-2400 Mhz)	Higher frequency bands (2-8 Ghz)
bandwidth	5-20 Mhz	100 Mhz(or more)
Switching design basis	Circuit and packet	All digital with packetized voice
Access technologies	W-CDMA, 1XRTT, Edge	OFDM and MC-CDMA(Multi carrier CDMA)
Component design	Optimized antenna design, multi-band adapters	Smarter antennas, multiband and wideband radios
IP	A number of air link protocols, including IP 5.0	All IP(IP6.0)

Fig. 4. Comparison of 3G and 4G.

5. BARRIERS IN PROGRESS

1. **Nobody is making the conversion to 4G:** All keep upgrading the present services i.e. 2.5G and 3G services as it is easy to upgrade the existing technologies because the equipment required are

already developed and hence is cheap to upgrade and hence nobody is making conversion to 4g as costly.

2. **Only few take up 4G:** As everybody doesn't switches to 4G hence its equipment will be costly and if few take up the conversion to 4G converters will be able to sell more services to their customers, it will not be enough to cover the higher costs of converting to 4G.

Providers have to plan carefully to make sure the cost is not too much.

One of technique being implemented in Asian networks in order to keep the cost realistic is the "Pay-Per-Use model of service"[5].

6. APPLICATION OF FOURTH GENERATION (4G)

1. Application of 4G based system is location **based service** [3]. It would be based mainly on visualized, virtual navigation schemes. These schemes support a remote database which would consist of graphical representation of streets and other physical characteristics of a metropolitan area. This data base could be accessed by a subscriber in a moving vehicle. And one would be able to see the internal layout or virtual representation of the environment ahead.

This is known as "**Telegeoprocessing**". Telegeoprocessing is a combination of Geographical Information System (GIS) and Global Positioning System (GPS). It will make it possible for public safety community to have wireless operational functionality.

2. In case of "**crisis-management application**"[3]: When natural calamities takes place and all the networks fail then restoring of communication quickly is done with wideband wireless mobile communication. With its help communication facilities like internet and video services can be set up within hours. Otherwise it would have taken days or weeks to set up communication in case of wired networks.
3. **Virtual Navigation:** As one is able to see the internal layout beforehand hence one can pre-plan security measures in case of any emergency as he can foresee circumstances. Example: In case of a kidnap the policemen will be able to see where the kidnapped person is kept and plan further steps how to go about rescuing the kidnapped people.

4. **Telemedicine:** In case of an emergency a doctor can set up a video conference with the specialist in order to get his assistance to save the victim.

5. Other applications are:

- Mobile IPTTV
- Social Networking services/user-generated content
- Mobile marketing and advertising
- Telematics
- Wireless VoIP Apps
- E-readers Apps/phones
- M2M Apps

7. LIMITATIONS FOR FOURTH GENERATION (4G)

Although it has so many advantages, still 4G has some limitations which are as follows:

1. There are many rural areas and buildings which are still not being served well by the current networks. The limitation of today's network will carry on to the next generations. 3G network has created unrealistic expectation of always on, always available, anywhere, anytime communication. Steps should be taken in order to correct the perception issue because later there will be a great deal of disappointment associated with the deployment of 3G and later 4G technologies and perceptions could become negative and if this happens none of the technologies neither 3G nor 4G will realize its full potential.
2. Another limitation is the cost to implement the next generation. The equipments required to implement the

next generation is expensive hence the implementation of next generation remains costly.

8. CONCLUSION

With time we can hope that 4G will be able to fulfill all the expectation of the user and all that it itself promises. It can also be said that the technological advancements are being made on a daily basis. These advancements will make a high speed data/voice over Internet-protocol (VoIP) Networks a reality.

It is also important that the industry must ensure that expectations are realistic and the services are able to meet the expectation.

The step by step evolution in wireless communication will give general public as well as the public safety community such amazing functionalities from the convenience of a single handled device.

REFERENCES

- [1] Andrea goldsmith "Wireless Communication" in 2005.
- [2] Pradipta Dasgupta, "E mergence of 3G wireless networks" in technology April 15, 2009.
- [3] <http://www.mobile.com>
- [4] <http://en.wikipedia.org/wiki/4g>
- [5] Jawad Ahmad, Ben Garrison, Jim Gruen, Chris Kelly and Hunter pankey "4G Wireless system" May 2, 2003.
- [6] Siti nor Faizah Binti Ab Malek "Evolution from 3G to 4G and beyond 5G".

5G Technologies

Vinay Kumar Singh¹, Shekhar Singh², Rachit Manchanda³, Shilpa Thakur⁴

¹M. E. Student,

^{1,2,3,4}NITTTR, Chandigarh, India

¹aspvinayrajput@gmail.com, ²shekariet@yahoo.in, ³rachit.tech2@gmail.com, ⁴shilpa.thakur75@gmail.com

Abstract: In this paper we present the concept of 5G Technology. 5G Technology is the Next Generation of mobile communication system. Presently 5G is not officially used because researches are going on it. In this research paper we will see personal view on 5G technology.

Keywords: Nano-technology, Cloud computing, All IP Network (all IP platform).

1. INTRODUCTION

In the present scenario we are having different wireless and mobile technologies in the field of communication which are distributed systematically like 3G and 4G wireless and mobile technologies. Now-a-days we are rarely using landlines because instead of using landlines we are preferring Mobile phones. Mobile phone not only keeps us connected to all over the world but also provide entertainment. In the world of telecommunication a lot of improvement from 1G to 2G and 3G to 5G.

5G technology is on the way to provide interface to most of the user that they can access their handset. Customers are aware of upcoming technology due to its affordable packages and its looks. Most of the developed countries are utilizing 4G technology and now they are trying to imagine that the whole technology will drawn in 5G like speed dialing, largest memory, audio and video players, Microsoft office, Bluetooth technology and all.

The 5G technology is a new revolution which is about to begin. The worldwide mobile phone will strike the surrounding who may call retrieve from china to Germany's local phone with this technology. 5G technology is going to be tough for laptop and normal computer. The present market will not be able to grasp this technology with mobile currently equipped with (GB) storage and latest operating system. This technology has a lot of features but under pipeline.

2. 5G NETWORK

The concept used by 5G network is flat IP because it is easier for different RAM and esteems in a single Nano-core network over 5G technology use nanotechnology Safety tools for security purpose that is due to the flat IP. Flat IP

network is the main concept to become 5G technology acceptable for all technology to meet customer criteria for data application which is being delivered over broadband network. Flat IP is tuned by network operator. Flat IP architecture is the way by which we can get the device using symbolic means, unlike normal IP used by hierarchical architecture.

- Mobile operator can take advantage by using Flat IP Architecture.
- For lower operation cost and capital expenditure minimize the network element in the data path.
- For the best service innovation towards the mobile and IP access network, we develop a flexible core network.
- Creating a roadmap for this will activate mobile broadband operator to make competition an perspective of price and performance.

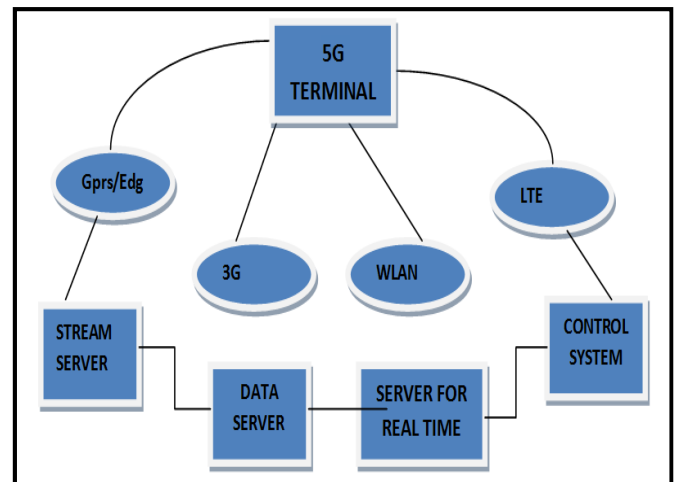


Fig. 1. 5G Terminals

3. 5G ARCHITECTURE -THE NANOCORE

5G NanoCore moves towards the technology which is listed below:

1. Nanotechnology
2. Cloud computing
3. All IP Platform

1. Nanotechnology

Nanotechnology is based on nanoscience to control the operation on nanometer scale i.e. 0.1 and 100 nm. This is also known as molecular Nanotechnology which control the overall structure of a matter depends on atom by atom and molecule by molecule engineering. The word Nanotechnology was first introduced in 1974 in an International conference at Tokyo.

Nanotechnology is being used for industrial revolution and no telecommunication industries are also using this technology in last few year. Nanotechnology shows its impact on mobile as well as core network in present era nanotechnology become very important part of telecommunication.

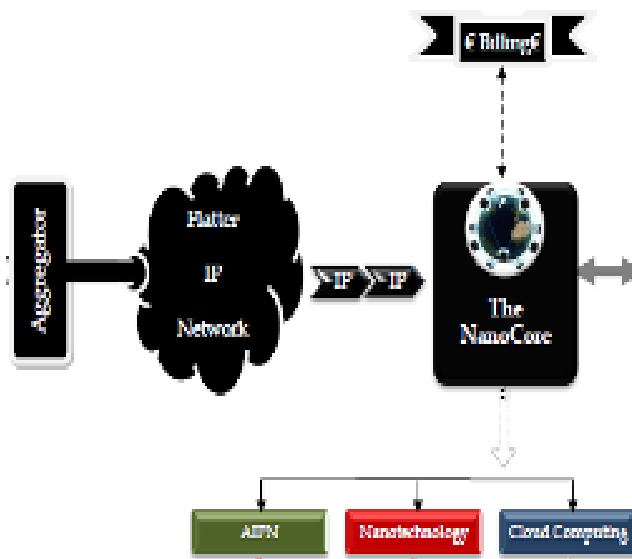


Fig. 2: 5G Architecture – Technologies under 5G NanoCore.

Nano-equipment

In modern world mobile phone become more than communication device which is turn into an individual identity. In Nanotechnology, mobiles are referred as Nano-equipments. The main focus of the wireless industry goal at ambient intelligence. Communication and computation should be available ready to provide interface to the user in intelligent way so it require a device as mobile together with intelligence to FIX deeply in human environment. Office, home and other public places that will enable computing communication and sensing.

1.2 Nano-equipment Specs

1. Transparent
2. Flexible
3. Self powered
4. Self cleaning

2. Cloud Computing

In telecommunication, technology uses the internet and main remote server to maintain data and application. In 5G network main, remote server works as our content provider. Cloud computing allows customer and business to use an application in absence of installation and get their personal life at any of the computer by using an internet. In nanocore we use same concept. With the help of nanocore use can tries to get his private account from data provider through nanocore.

Cloud computing provide operator with great opportunities. It also shows the importance of network and network development. It also need reliable and secure service providers operator should expertise in.

Operator can take any of cloud computing market and create new services. This technology allow our user to obtain real time application. To use this 5G network effectively and with the help of quantum cryptography, safe and reliable service can be provided.

Cloud computing has main three parts

1. Application: Demand based software service.
2. Platform: It is the second part of cloud computing. It explains which product we can use to deploy internet.
3. Infrastructure: This is the third part of the cloud computing which is known as infrastructure in the spine of the whole system.

To satisfy customer demand, 5G Nanocore uses all three parts which are listed above.

3. All IP Network

All IP network is last but not the least as we have already discussed for move towards different technology to a single 5G Nanocore.

We need a common path to communicate so Flat IP Architecture is necessary part of 5G network so due to increasing demand of mobile telecommunication all IP network is evolution of 3GPP system to meet consumer demand for real time applications. It is important from the performance and cost effective point of view.

Here some of the key point of Flat Architecture are:

1. Lower cost
2. Universal seamless access
3. Improve user experience
4. Reduce system latency
5. Decoupled radio access and core network evolution.

4. CONCLUSION

The Globalization of the wireless network and the mobile is using higher data rates and all IP principle presently. There are so many ways to radio access technology which provide IP based communication on the different network. Till today 5G is not being used as official term by any of telecommunication company such as Wi-max, 3GPP, and ITU-R for 5G. we have technology like Nanotechnology, cloud computing, all IP network.

5G include the entire feature which should be in telecommunication system for huge use in future.

5. CASE STUDY

Now we are curious to know that when this technology will be evolving, researches expect it by 2020. Still worldwide researches are going on all over the world for this technology.

REFERENCES

- [1] Toni Janevski, 5G Mobile Phone Concept, Consumer Communications and Networking Conference, 2009 6th IEEE.
- [2] Vasavi Bande, Mounika Marepalli, Leepika Gudur "Evolution of 4G-Research Directions Towards Fourth Generation Wireless Communication", " International Journal of Computer Science and Information Technologies", Vol. 2 (), 2011, 1087-1095.
- [3] B. G. Evans and K. Baughan, "Visions of 4G, " Electronics and Communication Engineering Journal, Dec. 2002.
- [4] H. Huomo, Nokia, "Fourth Generation Mobile, " presented at ACTS Mobile Summit99, Sorrento, Italy, June 1999.
- [5] J. M. Pereira, "Fourth Generation: Now, It Is Personal, " Proceedings of the 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, September 2000.
- [6] <http://www.4gwirelessjobs.com/pdf/5g-Wireless-architecture.pdf>
- [7] <http://www.5-g.co.uk/>
- [8] <http://5ginfo.blogspot.com/>
- [9] <http://www.beyond4g.org/wp-content/uploads/2011/03/5G-The-NanoCore.pdf> <http://www.beyond4g.org/vision-of-5g-networks-andarchitecture>.
- [10] 5G WIRELESS ARCHITECTURE-2010" By Vadan Mehta.
- [11] Amos Edward Joel (Bell Labs), "Cellular Mobile Communication System."
- [12] Andrew McGirr, Barry Cassidy (Novatel), 1992, "Radio telephone using received signal strength in controlling transmission power".
- [13] Douglas Fougnyes et al. (Freedom Wireless) 1998, "Security cellular telecommunications system".
- [14] Siegmund M. Redl, Matthias K. Weber, Malcolm W. Oliphant (March 1995): "An Introduction to GSM".
- [15] 5G Wireless Technology: 1.Manish Mathur, 2.Naresh Mathur, 3.Sumit Kumar, 4.Bhagyashree1, 2Department of Computer Science, Shekhawati institute of Engineering &Technology, Sikar, Rajasthan, India, 3 M.Tech,
- [16] Department of Computer Science and Engineering, JAGANNATH UNIVERSITY, JAIPUR, 4B.Tech THIRD YEAR, Department of Computer Science and Engineering, Poornma Group of Institutins, Jaipur.
- [17] <http://freewimaxinfo.com/5g-technology.html>.
- [18] <http://www.scribd.com/doc/22050811/5g-Wireless-Architecture-v-1..en.wikipedia.org/wiki/5G>.
- [19] <http://kevin-peter.hubpages.com/hub/3G-and-4G-Mobile-Services>.
- [20] <http://www.teknocrat.com/1g-vs-2g-vs-3g-vs-4g-vs-5g-comparison-differences-and-analysis.html>.
- [21] <http://www.ijcaonline.org/volume5/number4/pxc3871282.pdf>.
- [22] <http://www.globalreviewchannel.com/forum/3290-G-way.aspx.->

Role of Telecommunications Network in Universities

C.G. Nayak¹, Balbir Singh²

¹Department of Instrumentation and control Engineering

²Aeronautical and Automobile Engineering,

^{1,2}MIT, Manipal University Manipal, -576104, India

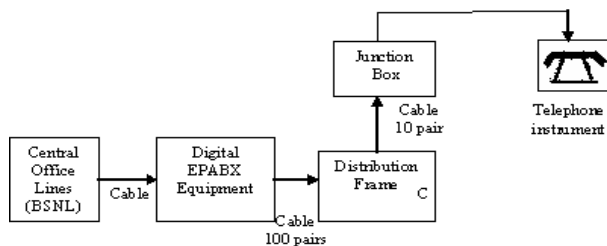
¹cgurudasnayak@yahoo.co.in, ²Balbir.S@manipal.edu

Abstract:—The wired digital Electronic Private Automatic Branch Exchange System (EPABX) is a telecommunications network is capable of handling voice, data, text, facsimile and image communication today. It is used in offices, industries, hostels, hotels, universities and in commercial complexes-networking or in transmission. However, during the study of wired digital EPABX system, noise problem is encountered. Therefore a new design is proposed that reduces the Noise in digital wired EPABX system using the DECT (Digital Enhanced Cordless Telecommunications) cordless technology. DECT is a WLL (Wireless Local Loop) system, based on the DECT standard. The DECT offers high quality access technology recognized by more and more users, regulators, standardization bodies, network operators and equipment manufacturers. It has proven multiple applicability as a network access in residential, offices, business and public environments showing easy mobility, speech quality comparable to wireless telephony. This research paper explains how the Noise can be reduced in wired digital EPABX System.

IndexTerms: DECT, MDF and WLL

1. INTRODUCTION

A typical Digital EPABX system architecture comprises C.O. Trunk lines, EPABX exchange, MDF Unit, Junction Boxes [1] and Telephone Instruments as shown in Fig: 1



Digital EPABX system
(PCM technology)

Fig. 1. Block diagram of Digital EPABX system

In this architecture when an incoming call originates from the central office exchange, a signal is generated by pabx trunk card in the epabx system [2] and then the signal is passed to MDF, junction box and telephone instruments through the copper cables. To feed a 100 telephone connections, we need 100 pairs underground or overhead copper cables. Furthermore, to give 10 telephone connections from junction box to the subscriber premises we need 10 pairs copper cables running from MDF to junction box [3]. Telephone instrument is connected to the junction box by single pair copper cable. Compact copper cable transmission lines suffer from the inherent problem of induced fields around them which results in noise. Noise is a phenomenon by which a signal is transmitted on one circuit or channel of a transmission system creates an undesired effect in another circuit or channel. In telecommunications or telephony, noise is often distinguishable as pieces of speech or signaling tones leaking from other people's connections. Hence, noise is a dominant factor in the performance of the above mentioned systems.

2. ANALYSIS OF THE WIRED DIGITAL EPABX SYSTEMS

The research work involving wired digital EPABX systems has been carried out at Telecom centers in Manipal and Mangalore, Karnataka, India. EPABX systems from Nortel Electronic Automatic Exchange (NEAX) 7400 ICS (Integrated Communications System) are available throughout the world. The proposed models for the accomplishment of the research work that are put into testing are NEAX 7400 ICS 140, NEAX 7400 ICS 150, NEAX 7400 ICS 160 and NEAX 7400 ICS 180 [4]. Mentioned models are tested for extension lines ranging from 100 to 1000 comprising parameter namely noise. The dissertation aims at making a comparative study of the above listed models with the proposed wireless digital EPBAX system.

Considering the calculation of noise, two attendant consoles are taken into account for all the models. If the disturbance is encountered while making a call from an operator console to a specific extension number, then it leads to noise. The

process is repeated for various extension numbers ranging from 2000 to 2999 depending upon the number of lines considered for testing. Each time the noise phenomenon is noted down relative to number of lines tested. We attempt to define noise in the following way: Number of lines encountering noise to the number of lines tested is expressed as percentage (%) of noise.

$$\% \text{Noise} = \frac{\text{No. of lines encountering Noise} * 100}{\text{No. of lines tested}}$$

In the testing of 300 line extensions in NEAX 7400 ICS 140 model, we choose extension numbers ranging from 2000 to 2299 and the percentage of noise obtained is 7 %. % noise = $(21/300) \times 100 = 7\%$ Where, No. of lines encountering noise = 21, No. of lines tested = 300.

Comparative Study of Noise for all NEAX models is shown in Figure 2.

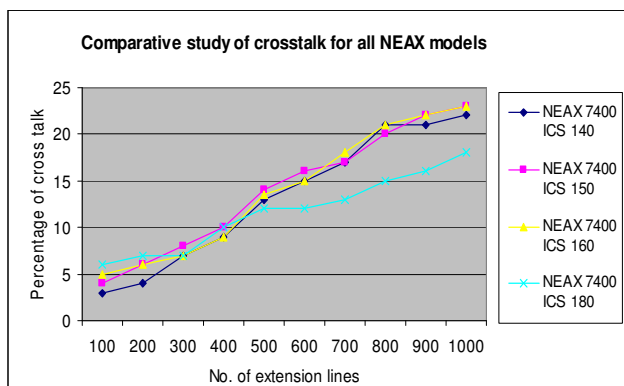


Fig. 2. Comparative Study of Noise for all NEAX models.

Percentage of Noise for all NEAX models is shown in Table 1.

Table 1. Percentage of Noise for all NEAX models

No. of lines tested	NEAX 140	NEAX 150	NEAX 160	NEAX 180
100	3	4	5	6
200	4	6	6	7
300	7	8	7	7
400	9	10	9	10
500	13	14	13	12
600	15	16	15	12
700	17	17	18	13
800	21	20	21	15
900	21	22	22	16
1000	22	23	23	18

3. DESIGN OF WIRELESS ARCHITECTURE FOR DIGITAL EPABX SYSTEM

A new design of the wireless digital EPABX system is proposed, which interfaces the DIU unit into the EPABX system through E1 Link and uses a Base Station instead of Junction Box. This reduces the use of copper cables when a link is established between the EPABX equipment and the Handset via DIU and Base Station. The proposed wireless architecture for digital EPABX system is shown in Figure 3.

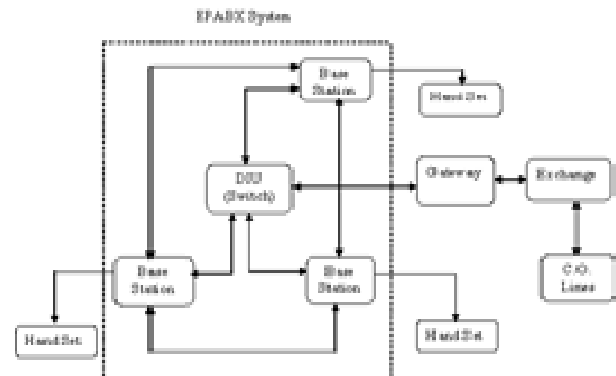


Fig. 3. Wireless Architecture for digital EPABX System

The various sub units of Wireless Digital EPABX System are Central Office Line, EPABX Exchange, DIU (DECT Interface Unit), E1 Link, BS (Base Station) and Handset [5] . In the new architecture of digital EPABX system C.O. Lines and Exchange remain the same as in wired EPABX system. Typically, with 4 E1 lines, the system can cater to about 1000 subscribers with 1:8 concentrations. It switches voice traffic to the telephone network using the V5.2 protocol to connect to an exchange. Frequency band used is 1880- 1900 MHz. The BS is a small, weather-proof and pole or wall mounted unit which is connected to and remotely powered from the DIU through three pairs of copper wires. Each pair carries 144 kbps data and power feed. One BS serves 50 subscribers in its neighborhood. The BS provides the radio interface between DIU and Hand Set. It is connected to the DIU through 3 pairs of twisted pair copper wires. The DIU feeds both power and signal. A DIU can be connected to up to 20 BS and each BS supports up to 12 simultaneous voice calls [6]. The handset is a small lightweight portable unit operated from rechargeable batteries. It allows the user to make calls from within the coverage area of any of the BS's connected to the DIU. The Handset has intelligence to handover seamlessly from one BS to another. The same HS can be used with different DIU's by appropriate re-registration when moving from one location to another.

4. ANALYSIS OF THE DIGITAL WIRELESS EPABX SYSTEM

An exhaustive study of constituent components in the designed wireless digital EPABX system is carried out at the Telecom Center, Mangalore Karnataka, India with utmost attention going into the parameters which have a direct influence on the performance of the system. The results obtained are compared judiciously with already existing design so as to get an idea of the profitability of the system. We have used NEAX (Nortel Electronic Automatic Exchange) 7400 ICS140 EPABX system for our testing purpose in the network so designed. The model is tested for same range of extension lines as earlier, so as to give a direct indicative comparison between the wired and wireless designs. Considering the calculation part in the testing of 100 line extensions in NEAX 7400 ICS 140 model, we choose extension numbers ranging from 2000 to 2099 and the percentage of noise obtained is 1 %.

$$\% \text{ noise} = (1/100) \times 100 = 1 \%$$

Where, No. of lines encountering noise = 1,
No. of lines tested = 100.

Study of noise in NEAX 7400ICS 140 model is shown in Figure 4.

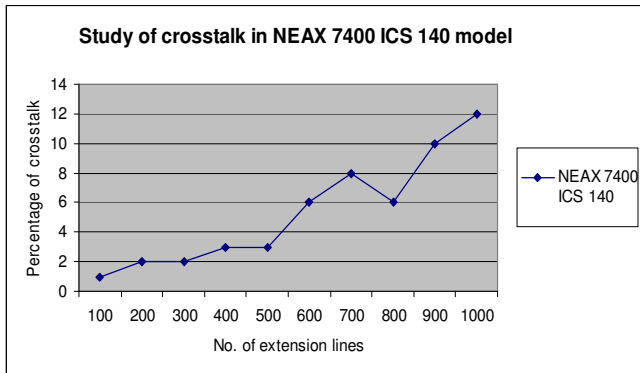


Fig. 4. Study of noise in NEAX 7400ICS model.

Percentage of noise in NEAX 140 model is shown in Table 2.

Table 2. Percentage of noise in NEAX 140 model

No of lines tested	NEAX 150
100	1
200	2
300	2
400	3
500	3
600	6
700	8
800	6
900	10
1000	12

5. PERFORMANCE EVALUATION OF WIRED AND WIRELESS DIGITAL EPABX SYSTEMS

The performance comparison of noise parameters for the wired digital EPABX system and wireless digital EPABX system using DECT cordless technology is shown through graphical analysis using bar graph. In reference to noise parameter shown in Figure 5, wireless digital EPABX system using DECT cordless technology has significantly less noise compared to wired digital EPABX system. The electromagnetic induction noise is seen in wired systems than wireless systems. Thus the noise can be reduced in wireless system. This has been proved in the research by testing various models of NEAX for 100 to 1000 lines. In our design, the incoming call generating from exchange to handset includes wireless connection from base station to handset. This design henceforth reduces noise to large extent as compared to wired digital EPABX system. From the bar graph analysis in Figure 5, noise is not completely removed. However, there is a significant reduction in the mentioned parameter to a great extent. Experiment has been carried out to measure the noise in wired 1000 extension lines exchange (NEAX 7400 ICS-140) and wireless 1000 extension lines exchange (NEAX 7400 ICS-140). It is found that in the wired exchange the noise varies from 3%- 22%, while the noise varies 1%- 12% in the wireless exchange. Thus the minimum reduction is 2 % in the case of 100 extension lines while 15% in the case of 800 extension lines as shown in figure 5. Hence it can be concluded that from our experiment the noise reduction varies from 2% - 15 % from wired system to wireless system.

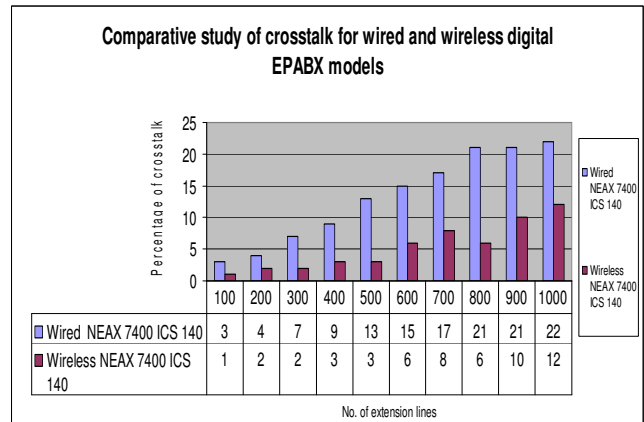


Fig. 5. Comparative Study of Noise for wired and wireless digital EPABX NEAX 140 ICS models.

6. CONCLUSIONS

This research paper involves an extensive study of the wired digital EPABX system a telecommunication network used in universities. The drawbacks are taken into account and a solution is proposed by a wireless design using DECT

technology. The detail study of the DECT system along with the designs of the various sub-systems of DECT established the ascendancy of the wireless design. The parameter like noise is compared with that of wired digital EPABX network. The percentage of noise is reduced to 2% - 15% depending upon number of extension lines tested for a range of 100 to 1000 lines for NEAX7400 ICS-140 models.

REFERENCES

- [1] Arlington, VA, (2004), "Telecommunications: multiline terminal systems: Requirements for PBX switching equipment, Addendum 1." by Telecommunications Industry Association, Standards and Technology Department. USA.
- [2] Lawrence Harte (2005), "Introduction to Telecom signaling," 2nd Edition.
- [3] Dr. P. N. Das (1989), An Introduction to Automatic Telephony, 8th edn., Modern Book Agency Private LTD., Kolkata
- [4] NEAX 7400 ICS Integrated Communications System, General Description Model 140/150/160/180, Issue 1, June 2008, NEC Corporation, Japan.
- [5] CorDECT Wireless Access System, December 2000, Midas Communication Technologies Private Limited, Chennai.
- [6] Lan Poole, (2010), "Cellular communications explained: from basics to 3G," by Newnes, England.

A New More Efficient, Dynamic, and Robust Access Control Scheme over Wireless Sensor Networks

Raj Kumar¹, Ritesh Kumar²

¹Computer Science and Engineering, NIT Hamirpur, Hamirpur, India
nit.er.raj@gmail.com

²Electronics and Communication Engineering, P.K.I.T.M. Mathura, Mathura, India
er.ritesh29@gmail.com

Abstract: Owing to the characteristics of sensor devices, they are easily compromised by an adversary, who inserts misleading informations or modifies transmitted messages to devastate the whole sensor networks. On the other hand, nodes in the sensor networks may be lost because of power exhaustion or malicious attacks. So, entire sensor networks would be ended after some time of operation.

To extend the life time of sensor networks, new node should be necessarily deployed. To prevent the malicious nodes from joining the sensor networks, an access control is a designed requirement for controlling the deployment of new sensor node. Based on the secure one way hash function, this paper presents a new dynamic access control scheme over wireless sensor networks (WSNs). The proposed scheme not only prevents malicious nodes from joining sensor networks, but also key establishment is included in the authentication procedure. Compare to the authentication and key establishment procedure of previously proposed schemes, our scheme provides very simple and efficient procedure. Along with, it could offer computational efficiency, energy, and bandwidth savings in a great extent.

IndexTerms: Wireless Sensor Networks (WSNs), Access Control, ECC, and ECDHP.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) is the most important and suitable technology to monitor challenging hostile environmental events like wildlife monitoring, military sensing and tracking, distributed measurement of seismic activity etc. To save manufacturing cost, a sensor node is usually built as a small device, which has limited memory, a low-end processor, and is powered by a battery [16]. The position of sensor nodes cannot be engineered or predetermined in sensor networks.

And, because of constraints of the sensor devices, after several days or weeks of operation, some nodes in the network may exhaust their power or lost [2]. That is, nodes in a sensor network may exhaust their battery or be stolen by an adversary. In addition, some nodes may be destroyed by adversaries so that the entire network may become useless.

To extend the lifetime of the sensor networks, new node deployment is necessary. However, an adversary can also deploy malicious nodes into the network for eavesdropping purposes or inserting false reports. Hence, a new node should prove that it is a legitimate node through the authentication [8]. For the authentication of legitimate sensor node, it requires some classical techniques like Puzzle solving done by sensor nodes (New deployed sensor node or old deployed node) with already installed key materials (given by Base Station), mechanism of exchanging some already installed keys with a little bit of computation and less communication during authentication, exchanging signature given by Base Station, and or an involvement of Third party (Trust Center) etc. Also, to protect floating messages in network, it requires secret key to encrypt those messages.

Today's, we have a lot of approaches for the distribution of keys in sensor nodes but most practical approach is key pre distribution. In this approach, keys are pre installed in sensor nodes and make secure communication among the sensor nodes using procedures like, nodes having common keys to make secure communication, or making secure communication through this master secret key etc. But, all of those approaches are introduced above, would not provide network resiliency. Recently, many key predistribution schemes were proposed to protect sensor networks [3–4]. Here, we are trying to use such scheme that provides mainly local communication with simple operation, energy efficient, and pairwise in nature.

To prevent malicious nodes from joining our sensor network, an access control is required in the design of new node insertion phase. Some researchers tried to detect malicious nodes after deployment in the sensor networks; however, an adversary harms the sensor networks. This paper proposes a more efficient, dynamic and robust access control in the sensor networks using secure one way hash function. Elliptic Curve Cryptography (ECC) and Diffie-Hellman algorithm over elliptic curve (ECDHP) are used during authentication and key establishment procedure to provide same security level with a less number of bits

(Compare to RSA). Here, we introduced the concept of expiration time, that made this scheme more robust in future time (no impact of previous compromised node). With simple calculation, few communication overheads and less memory uses, make this scheme more practical towards real life implementation of sensor networks. Moreover, this scheme can be easily implemented as dynamic access control scheme as all the old secret keys and broadcasting information of existing nodes are not updated once a new node is added. Also, under the security of secure one way hash function, ECDHP and ECC, it would try to provide more security aspects.

The rest of this paper is organized as follows. In Section 2, we present necessary background in terms of security threats and discussion of related works. In Section 3, we present the proposed scheme. In Section 4, discussions are done in terms of security measurements, performances and comparison with the related work. In the last Section, some conclusion is made.

2. BACKGROUND

A. Review of Attacks

It is seen that generally a Wireless Sensor Network (WSNs) is setup in a hostile area i.e. having lack of infrastructures.

Table 1: NOTATION

Symbol	Description
Q	A large prime number
E	Elliptic Curve
F_q	A finite field over prime number
E_q	Elliptic curve over finite field F_q
P	A point over E_q A generator of group G
Z_q	A set of large prime numbers
X	A random number from Z_q
N_i	An identity of the node N_i
k_i	A secret key of the node N_i
T_i	Expiration time of node N_i
$h()$	A secure one way hash function
PLUS	+ operation
MINUS	- operation
T	Broadcasted Time by System (Base station)
K_{ij}	A session key established between N_i and N_j

So, there is a possibility of vulnerabilities by an adversary in the sensor networks. The adversary can directly deploy malicious nodes to eavesdrop messages or modifies the floating messages and finally disrupts the network functionalities in the sensor networks. At first, Sybil Attack, a particularly harmful attack in sensor networks where a malicious node behaves as if it were a larger number of nodes [15], for example by impersonating other nodes or simply by claiming false identities. The malicious node may

be deployed directly by adversaries or just a compromised one. The pictorial view of Sybil attack is shown in the Figure 1(a), where an adversary node AD behaves like more than one node. From the perspective of node A, node AD appears as node D, from the perspective of node C, as node B etc. It is very dangerous to the sensor networks. And, the wormhole attack, that makes a tunnel in the sensor networks by two nodes (either compromised legitimate node or malicious node deployed by attacker) and produces great concerns in the sensor networks by making two remote nodes as much nearer to each other as depicted in figure 1(b). It produces great harms to the routing protocols in the sensor networks as well as a challenge to the whole system. It confuses the entire sensor networks. Now, node replication, it is a direct replication of the legitimate nodes throughout the sensor networks. It makes a lot of clones of a single or a set of compromised legitimate nodes. It is shown in figure 1(c). A lot of other security threats are present in the sensor networks but only a few are discussed here.

B. Related Work

A lot of work has been already done in the era of access control in WSNs. Some researchers focused either on much more security aspects or on easy implementation. Zhou et al.'s [8] scheme is a first dynamic access control which focused on a greater security provision with use of bootstrapping time. But the use of bootstrapping time made it impractical for real life implementation. Also, its security claims come under suspicion, due to unavailability of checking bootstrapping-time for the existing old nodes [8]. Zhou's scheme also bears high computational and communication cost (because use of three multiplication and one inverse operation). Hui-feng Huang [10] proposed a dynamic access control protocol based on Elliptic

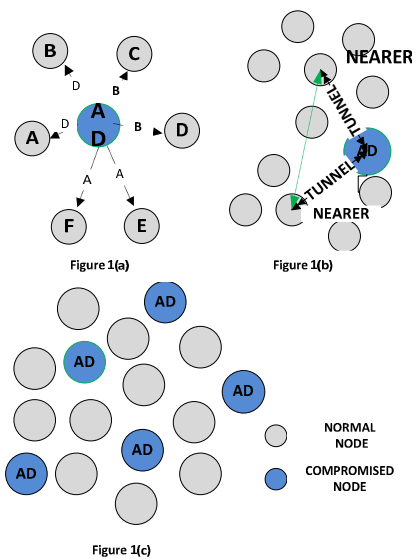


Fig. 1.

Curve Cryptography (ECC), and also considering the real life implementation of access control protocol in Wireless Sensor Networks (WSNs) [8]. He used cascaded hash chain as a main weapon in authentication and key establishment phase to make this protocol practically possible. Also, he strengthened security measures but, unfortunately it did not perform as expected and its security holes are caught by other researcher. H.F. Huang proposed a more secure and efficient Access Control Protocol in WSNs based on ECC and the concept of Schnorr signature [2]. Basically, it exploited the concept of time bound in which once the time period elapsed the sensor node in wireless networks cannot access any data for further time period. Actually this thing was done, in keeping the mind that we have to reduce long time impact of the node compromised problem. Also, this proposed scheme presented a simple and efficient with a little more computational cost in authentication and key establishment phase, and tried to reduce overhead to a great extent. Hui-Feng Huang et al.'s scheme [9] tried to make it most efficient with offering few security measures as necessary and could co-exist them in real life implementation with least cost in terms of computation, communication and memory used. But, little bit of security measures makes this access control scheme not suitable in these days. This scheme only favors resource constraints issue of wireless Sensor Networks (WSNs). So, there is a basic need of such access control scheme that could claim a greater security measures with feasibility towards computational, communicational and storage requirements. In this paper work, we tried to use such operator which incur little computational cost, as well as tried to make whole thing as local without communication to the base station. We introduced the time bound concept in this scheme to make our protocol for future safety (no further impact of any compromised node). The cryptography tool is ECC same as other researchers to utilize the merits of ECC rather than RSA. Thus, we ultimately tried to make more efficient, dynamic and much robust access control scheme for a WSNs. We always kept in mind to make them more secure, efficient plus more practical towards real life implementation.

3. THE PROPOSED SCHEME

We assume that all sensor nodes have the same transmission range and can communicate with each other. At the beginning, a lot of sensor nodes equipped with all needed resources are deployed in the designated area. It is no doubt that after sometime, there is a need of new node deployment to extend the span of sensor networks, because of nodes in a sensor networks may be lost or destroyed. Also, we assume that span of our sensor networks is divided into t periods, numbered 1, 2, . . . t [2]. For simplicity, we let t be an integer, that is, the system needs new nodes deployment at the end of time t . This maximum number of (expired) time period t should not be considered as a limitation of the system.

Without loss of generality, the proposed method would accomplish two tasks [9].

1. Node Authentication

A deployed node establishes its identity with its neighboring nodes and shows that it has the right to access the sensor networks through authentication.

2. Key Establishment

In course of authentication, pairwise shared keys would be created between a deployed node and its neighboring nodes through handshaking of data to provide secure communication among them. This also needs some simple computation.

The proposed scheme is based on secure One-way Hash Function, Elliptic Curve Cryptography (ECC), and Diffie-Hellman algorithm over elliptic curve (ECDHP). The proposed scheme consists of three states: Initialization phase, Authentication and Key-establishment phase and New Node Addition phase. They are followings-

Initialization Phase:

Before a sensor networks is deployed, the system (Base Station) chooses a large prime number q ($q \approx 2^{160}$) and an Elliptic Curve E_q , a cyclic group $G = \langle P \rangle$ of points over the elliptic curve E_q , where P is the generator of the group and has an order n of at least 160 bits [5–7]. Then, the system (Base Station) selects a random number and generates secret keys for sensor nodes as $k = xP$ of the point over the elliptic curve E_q . Then, the system (Base Station) chooses a secure one way hash function $h()$.

Firstly, the Base Station preloads the Elliptic Curve E_q , the generator P of the group G over the elliptic curve E_q to each sensor node. Then, System (Base Station) chooses an expiration time T_i and generates a number of r secret keys $k_i = x_i P$ for $i=1,2,\dots,r$, and preloads each secret keys k_i , expiration time T_i and secure one way hash function $h()$ to the identity(ID) of node N_i for $i=1,2,\dots,r$. Next, the system (Base Station) computes $S_{ij} = h(k_i, T_i, N_j)$ PLUS $h(k_j, T_j, N_i)$ for $i=1,2,\dots, r$ and $j= i+1, i+2, \dots, r$, and broadcasts all information to the deployed nodes $\{N_1, N_2, \dots, N_r\}$.

Authentication and key establishment phase:

The process of authentication and key establishment for two nodes N_i and N_j is described in the following steps.

Step 1: The node N_i selects a random number t_i and computes $A_i = t_i P = (A_{xi}, A_{yi})$ over the elliptic curve E_q . Then node sends A_i, T_i with its identity N_i to another neighboring node.

Step 2: Other neighboring node N_j checks

If ($T_i > T$) **then** /* (T periodically broadcasted by Base station)

ACCEPT

And selects a random number t_j and computes $A_j = t_j P = (A_{xj}, A_{yj})$. Then computes shared session key $K_{ij} = t_j A_i = t_i t_j P = (K_{xij}, K_{yij})$. Computes $a_j = S_{ji}$ MINUS $h(k_j, T_j, N_i)$ using broadcasted data of S_{ji} and finally computes $Z_j = h(a_j, K_{ij})$. Sends Z_j, A_j, T_j with its identity N_j to neighboring node N_i .

Else

DISCARD. /* (Either illegal node or old legitimate node)

Step 3: Firstly, Node N_i checks

If ($T_j > T$) **then**

ACCEPT and then node N_i computes the shared session key $K_{ij} = t_i A_j = t_i t_j P = (K_{xij}, K_{yij})$. After receiving Z_j, T_j and using preloaded data

If ($h(h(k_i, T_i, N_j), K_{ij}) = Z_j$) **then**

The node N_i can make sure that node N_j is a legitimate node. N_i also computes $a_i = S_{ij}$ MINUS $h(k_i, T_i, N_j)$ using broadcasted data of S_{ij} and finally, computes $Z_i = h(a_i, K_{ij})$. Then it sends Z_i .

Else

DISCARD. /*(Not legitimate node)

Else

DISCARD /* (Either illegal node or old legitimate node)

Step 4:

Now Node N_j checks legitimacy

If ($h(h(k_j, T_j, N_i), K_{ij}) = Z_i$) **then**

Node N_j can make sure that node N_i is a legitimate node also.

Else

DISCARD. /*(Not legitimate node)

In this way, both nodes N_i and N_j make authentication to each other and make a shared common session key with handshake and simple mathematical calculation.

With the broadcasting information $S_{ij} = h(k_i, T_i, N_j)$ PLUS $h(k_j, T_j, N_i)$, it provides that $a_i = S_{ij}$ MINUS $h(k_i, T_i, N_j) = h(k_j, T_j, N_i)$ and $a_j = S_{ji}$ MINUS $h(k_j, T_j, N_i) = h(k_i, T_i, N_j)$. For the security reason to create a different session key for other pairing nodes, the random number t_1 and t_2 should be used only one time. The outline of above procedure is shown in figure2.

Proposed Scheme: Authentication & Key-Establishment

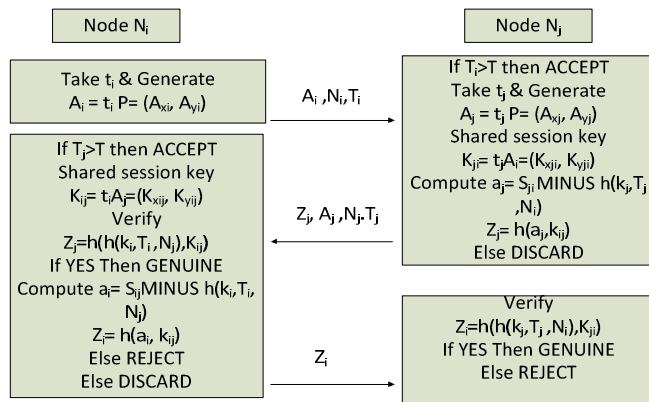


Figure 2

New Addition Phase:

During the network operation, if some sensor nodes are lost or destroyed, new sensor nodes are needed to be deployed. When a new node with identity N_{r+1} is added, the System (Base Station) first generates a secret key with a random number x_{r+1} as $k_{i+1} = x_{r+1} P$ and preloads the Elliptic Curve E_q , the generator P of the group $G = \langle P \rangle$ over the elliptic curve E_q , expiration time T_{r+1} and secure one way hash function $h()$ with secret key k_{i+1} . Next, the System (Base Station) only computes $S_{i(r+1)} = h(k_i, T_i, N_{(r+1)})$ PLUS $h(k_{(r+1)}, T_{r+1}, N_i)$ and broadcasts the new information $S_{i(r+1)}$ for $i=1, 2, \dots, r$, to inform its existing nodes. The authentication and key establishment for any old node with the new node N_{r+1} would be the same as mentioned above. The other informations in existing nodes are not updated. Hence, the proposed scheme can be easily implemented as a dynamic access control because all the old secret keys and broadcasting messages of existing nodes need not be changed once a new node is added or an old node is lost.

4. DISCUSSION

The proposed scheme is based on Secure One-way Hash Function, Diffie-Hellman algorithm over elliptic curve (ECDHP) and Elliptic Curve Cryptography (ECC). And, the security of our scheme is bounded in the security strength of these terms. Here, we reviewed the security measures of these terms needed for further security analysis [5–7, 12, 13].

Definition 1. A secure hash function, $h() : x \rightarrow y$, is one-way; if given x , it is easy to compute $h(x) = y$; however, given y , it is hard to compute $h^{-1}(y) = x$.

Table 2: Complexity of generic attacks on different properties of hash functions [11].

Property	Ideal security
One-wayness	$2^{(n-1)}$
Second preimage-resistance	$2^{(n-1)}$
Collision-resistance	$1.2 \cdot 2^{(n/2)}$

Definition 2. The elliptic curve discrete logarithm problem (ECDLP) in E_q is as follows: given $P \in E_q$ with order n (that is $nP = O$) and Q is a point in the cyclic group $G = \langle P \rangle$. It is intractable to find r such that $Q = rP$.

Definition 3. The elliptic curve computational Diffie-Hellman problem (ECDHP) is as follows: given $t_1 P$ and $t_2 P$ over elliptic curve E_q , it is hard to compute $t_1 t_2 P$ for any positive integers t_1 and t_2 .

A. Security Analysis

In the proposed scheme, at first there are r nodes which are deployed in the designated area. The System (Base Station) chooses an expiration time T_i and generates a number of r

secret keys $k_i = x_i P$ for $i=1,2,\dots,r$, and preloads each secret key k_i , expiration time T_i and secure one way hash function $h()$ to the identity(ID) of node N_i for $i=1,2,\dots,r$. Next, the system (Base Station) computes $S_{ij} = h(k_i, T_i, N_j)$ PLUS $h(k_j, T_j, N_i)$ for $i=1,2,\dots, r$ and $j= i+1,i+2,\dots,r$, and broadcasts all information to the deployed nodes $\{N_1, N_2, \dots, N_r\}$. The security of proposed scheme is based on security strength of secure one way hash function, ECC, and ECDLP. Therefore, even if an adversary gets the broadcasted information $S_{ij} = h(k_i, T_i, N_j)$ PLUS $h(k_j, T_j, N_i)$, there is no way to derive secret key (k), expiration time (T) of nodes, because of, these are protected under secure one way hash function $h()$. So, the adversary could not get secret key and expiration time exactly what are being used by a legitimate node. Thus, it can withstand some generic attacks. However, due to less temper resistant, if some nodes are compromised by an adversary then he can extract the secret keys and expiration time of those nodes and can only use these key materials for a specified period of time (until the expiration time) i.e. no future impact of that compromised node in the sensor networks. And, if an adversary tries to manipulate the existing expiration time (T) to extend the duration of expiration time then corresponding signature $S_{ij} = h(k_i, T_i, N_j)$ PLUS $h(k_j, T_j, N_i)$ would become invalid for further usage (for authentication and key establishment) and finally we cannot do such manipulation. Also, due to the uniqueness property of secret keys and expiration time, there is no possibility of pretending to someone (other legitimate node) else by an adversary. So, our scheme could also prevent node masquerading attack and Sybil attack, because, there is no way to produce such a secret key and expiration time which is used for more than one legitimate node simultaneously in the sensor networks.

And, this scheme also prevents new node masquerading attack, because of the new secret key and an appropriate expiration time which is only given by the Base Station at the time of deployment. So, there is no way to get new node's secret key and its expiration time by an adversary and launch new node masquerading attack either by new key materials or manipulating old (valid for a limited period of time) one to a new one. In this scheme, legitimate nodes make a pairwise shared common key using ECDHP and securely communicate using them. Thus, without knowing pairwise shared secret keys, an adversary cannot eavesdrop the transmitted packets or modify the floating packets in the sensor networks. An adversary could only get $A_i = t_i P$ and $A_j = t_j P$, but, it would be very hard for them to obtain t_i and t_j from it (as they are protected under the security of ECC). Thus, without knowing t_i and t_j , an adversary would obtain nothing about the secret shared session key K_{ij} . So an adversary can neither insert false messages into the sensor networks nor modify the transmitted messages in the sensor networks. Hence, the proposed scheme could withstand the Eavesdropping. Even if an adversary compromises any node (due to less temper resistant of sensor node) then, however,

it can only know about that compromised sensor node's secret key and pairwise shared keys. So, the effect of that compromised node would remain in the vicinity of that node and also, for a limited period of time (during expiration time only). So, the resiliency of the entire sensor networks would not be affected. In our scheme, once the expiration time has elapsed, an adversary could not do anything more in our sensor networks. So, our scheme could resist the wormhole attack (which refers that two compromised nodes makes a tunnel and send information to the other distant side and try to produce a false scenario that two remote nodes seem to be nearer and impairs the existing systems e.g. Routing Protocol etc.) by just exploiting the concept of expiration time of compromised nodes. Once, the expiration time expires, wormhole attack automatically finishes. On the other hand, if an adversary tries to manipulate expiration time of legitimate, then corresponding signature $S_{ij} = h(k_i, T_i, N_j) \oplus h(k_j, T_j, N_i)$, would become invalid for that legitimate node. So, the wormhole attack can be used only in a limited interval of time (during Expiration time only) and further network resiliency remains in good position. Our scheme could not stop the node replication attack completely but, it allows only node replication attack for a limited duration of time (Before the expiry of expiration time of the compromised legitimate node). Thus, our scheme provides a greater security provision compare to other.

B. Performances

The proposed scheme uses ECC as the cryptography tool. Compared with RSA, ECC can achieve the same level of security with smaller key sizes. It has been shown that 160 bits ECC provides comparable security to 2048 bits RSA [14]. Under the same security level, smaller key sizes of ECC offer merits of faster computational efficiency, as well as memory, energy and bandwidth savings [8].

C. Communication Overhead

Our scheme does not require any central trusted system or base station. There is no need of informing state of nodes to the central trusted system. Authentication and key-establishment is done locally by using the broadcasted informations of base station. Thus it would be more suitable and practical in terms of communication overhead.

D. Computation Cost

The proposed scheme needs only two multiplications over an elliptic curve and three one way hash function computation. This is much less computational cost than other scheme where a lot of computation cost incur in the computation of inverse operation, multiplication operation over elliptic curve, operation of hash function & hash chain etc. Thus, proposed scheme could be verified as efficient in computation.

Storage

This scheme does not demand a greater memory in keeping required data which is used during authentication and key establishment. So, it could be feasible in storage constraint.

C. Comparison with Related Work

For the comparison of related works, here, we have not discussed more about the related works (for more discussion go to background section 2.); only the main comparison is shown in the form of charts. For comparison with related work, let the notation T_m the time for one multiplication computation over an elliptic curve, T_i the time for one inverse computation, T_h the time for computing the adopted secure one way hash function [9]. Note that the times for computing modular addition and PLUS or MINUS operation are ignored, since they are much smaller than T_m , T_i , T_h . Moreover, a hashing computation is more efficient than T_m , T_i . The comparison chart of our proposed scheme with other existing schemes is given below in chart 1 and chart 2.

5. CONCLUSION

We focused on the prevailing weakness of previous schemes of access control in WSNs and found that they suffers either from lack of security aspects or much more costs (computation, communication & memory used). Here, we proposed an efficient scheme in terms of computational cost, communication overhead and memory use and in parallel way providing maximum security aspects. We used only less computational cost operation to make this protocol practical towards real life implementation. We added a little more cost but did not leave much more security holes. Also, we made whole scenario as local as possible so that it could appear as efficient in terms of communication overheads. To provide little and less memory use with same security strength, we used Elliptic Curve Cryptography instead of RSA. Thus, it provides both efficiency and well security measures.

Chart 1: Computation for each node to achieve authentication and compute a common key

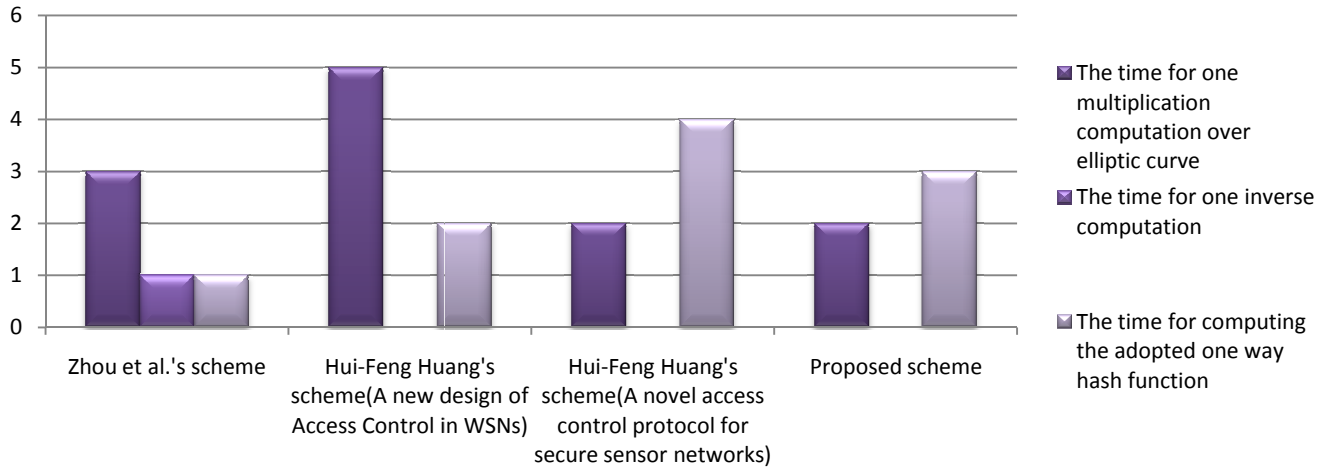
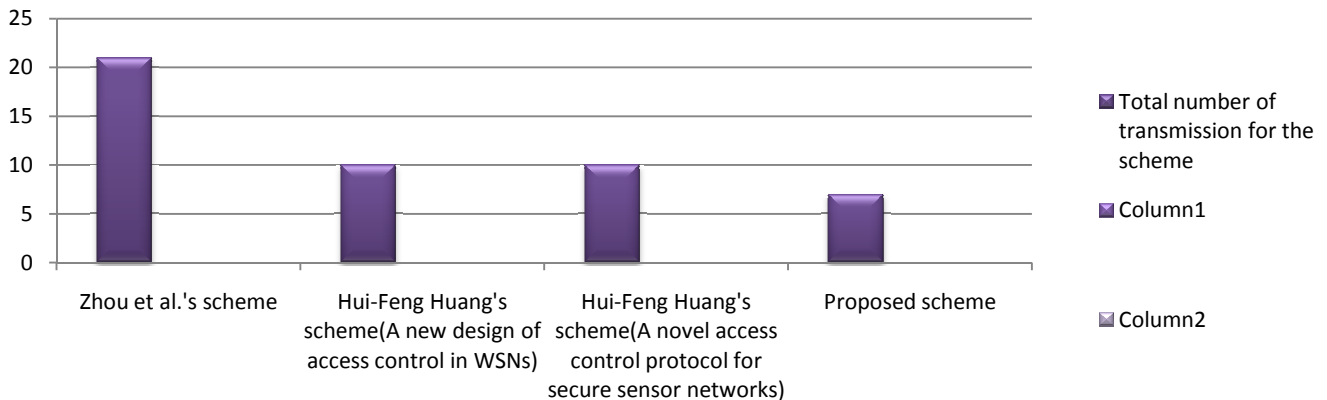


Chart 2: The total number of transmissions (communications) per node for the scheme



REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.
- [2] Hui-Feng Huang, "A New Design of Access Control in Wireless Sensor Networks" *International Journal of Distributed Sensor Networks* Volume 2011, Article ID 412146, 7 pages.
- [3] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proceedings of IEEE Symposium on Security and Privacy*, pp. 197–213, May 2003.
- [4] H. F. Huang, "A new design of efficient key pre-distribution scheme for secure wireless sensor networks," in *Proceedings of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP '07)*, pp. 253–256, Kaohsiung, Taiwan, November 2007.
- [5] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, pp. 203–209, 1987.
- [6] V. Miller, "Uses of elliptic curves in cryptography," in *Advances in Cryptology (CRYPTO '85)*, vol. 218 of *Lecture Notes in Computer Science*, pp. 417–426, Springer, 1986.
- [7] S. Vanstone, "Responses to NIST's proposal," *Communications of the ACM*, vol. 35, pp. 50–52, 1992.
- [8] Y. Zhou, Y. Zhang, and Y. Fang, "Access control in wireless sensor networks," *Ad Hoc Networks*, vol. 5, no. 1, pp. 3–13, 2007.
- [9] Hui-Feng Huang, Kuo-Ching Liu, "A New Dynamic Access Control in Wireless Sensor Networks," 2008 *IEEE Asia-Pacific Services Computing Conference*.
- [10] Hui-Feng Huang, "A novel access control protocol for secure sensor networks," *computer standards & interface* 31, 2009.
- [11] Ilya Mironov, "Hash functions: Theory, attacks, and applications," *Microsoft Research, Silicon Valley Campus*, November 14, 2005.
- [12] R. Kumanduri, *Number Theory with Computer Applications*, Prentice Hall, Upper Saddle River, NJ, USA, 1998.
- [13] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.
- [14] S. Vanstone, Responses to NIST's proposal, *Communications of the ACM* 35 (July) (1992) 50–52 (communicated by John Anderson).
- [15] James Newsome, Elaine Shi, Dawn Song, and Adrian Perrig, "The Sybil Attack in Sensor Networks: Analysis & Defenses," *IPSN'04*, April 26–27, 2004, Berkeley, California, USA, 2004.
- [16] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

Mobility Management Issues in Hierarchical Mobile IPv6 in 4G Networks

Shamurailatpam Susanta Sharma¹, Himanshu Sharma²

*School of computer sciences and Engineering
Lovely Professional University, Punjab, susantasharma031@gmail.com
Lovely Faculty of Tech. and Sciences
Lovely Professional University, Punjab
himanshu.16833@lpu.co.in*

Abstract - The future 4g networks will satisfy the user's need of mobility, information access and independence. Many firms are now working towards realizing 4G Networks. 4G networks are known for their smooth connectivity between existing networks which include GSM, UMTS, GPRS and Wireless LAN etc. 4G networks are the collection of heterogeneous environment with different access technologies that vary in bandwidth, handoff latency and cost. Seamless connectivity in such networks entirely depends on efficiently handoff mechanisms. In this paper we represent a handoff Management Mechanism based on Hierarchical Mobile IPv6 (HMIPv6) with the combination of session imitation protocol. The research discussed here, reflect seamless handoff management across different over the heterogeneous network

Keywords: 4G Networks, Quality of service, MIP, HMIP, VOIP, MIPv6, HMIPv6, Care of Address, Binding update, Triangle Routing, Mobility Management protocol, Vertical Handoff, Handoff mechanism, SIP.

1. INTRODUCTION

Mobile communication is developing very rapidly and with passages of time, new technologies are being introduced to facilitate the mobile users more from the technology. The past technologies are replaced by new ones and needs are growing for the new technologies to be developed. One such development is 4G Networks. The introduction of 4G has widened the scope of mobile communication. Now mobile is not only a device used for talking but it's more or less a portable computer that can serve different purpose. 4G offers higher data rate with seamless roaming. The mobile user can communicate without any disturbance while switching his coverage area.

The 4G has been developed with the aim of high data transmission speed as well as to accommodate quality of services (QoS) feature. 4G user will have more demands for seamless roaming across different wireless network, support of various services (e.g. multimedia application and QoS guaranteed). The mobile technologies will provide access to users, simultaneously providing high bandwidth, low latency, low power consumption and wide area data service

to a large number of mobile users. For this, we need to design a system that gives the benefit to the overall characteristics.

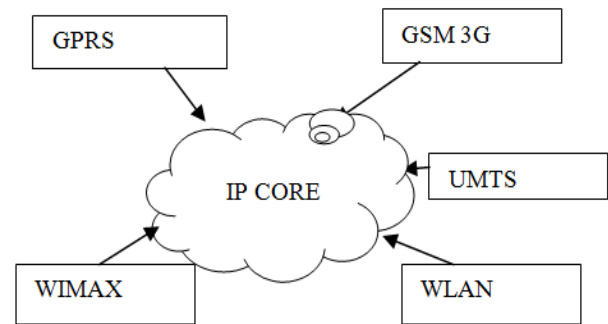


Fig 1. A sample of 4G Network

In 4G, Mobility management is a logical concept rather than a physical. It provides seamless mobility management support of the user while the handoff process is done. Resourceful Mobility Management would be a key area for 4G Networks. Handoff Management is an integral element of Mobility Management. In this paper, we provide a combination of Mobility Management Protocol i.e. Hierarchical Mobile IPv6 and Session Imitation Protocol (SIP) on 4G network during vertical handoff.

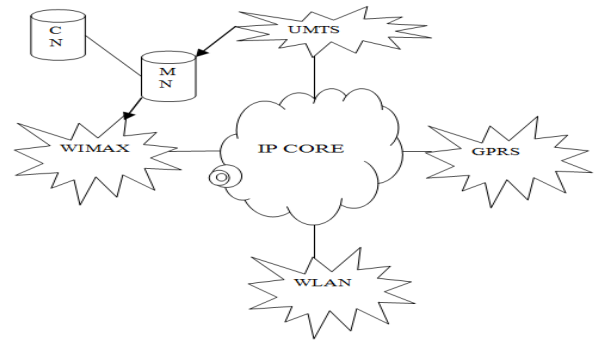


Fig. 2. A sample of 4G Architecture

1.1 Handoff

Traditionally handoff management stands for maintaining smooth and seamless communication while the Mobile Node (MN) moves inside or outside of the current network. MN can move inside one network Access point (AP) to another AP, while MN can also stay in its current coverage network. During both cases, MN undergoes handover process.

The handover inside MN coverage network is named as horizontal handover while the handover in which the MN changes network such as MN moves from GSM to UMTS, is called vertical handover. In the 4G Network the implementations of vertical handover is more challenging as compared to horizontal handover.

Mobility is one of the most emphasised requirements of communication in significantly advanced technology era. Mobile communication requires being mobile in real sense, i.e. to support multi heterogeneous while on move. That is only possible if there is some sort of correlation among these heterogeneous networks.

1.2 Types of Handoff

The mobile node should be there on the access point i.e. Base Station (BS) and their coverage areas are known as Cell. The cell size will depend on the type of network e.g. GPRS, WLAN, and UMTS etc, size of the BS and the transmission power of the BS. In term of the same network of the cell will be overlapped with each other. Also, in the 4G network the cell of different network may overlap with another network cell. This makes vertical handoff possible at any time. The handoff may occur on different factors like in 4G Networks. They contain heterogeneous network and may contain GPRS, WLAN, and UMTS etc. For this they have different characteristics like signal strength, load balancing, number of connection, frequencies.

- **Horizontal handoff:** The handover inside MN coverage network when mobile node changes to another node within the same network is named as horizontal handover.
- **Vertical handoff:** The handover, in which the MN changes network for example MN moves from GSM to UMTS, is called vertical handover.
- **Intracell handoff:** The mobile node changes the frequency for communication within same cell.
- **Intra frequency handoff:** The mobile node will move to next access point with same frequency.
- **Inter frequency handoff:** In GSM Network the mobile can move from one base station to another base station operating at different frequency

- **Hard handoff:** Occurs when the mobile node needs to disconnect the old base station before establishing the connection of new base station.
- **Soft handoff:** In CDMA, when the mobile node moves it can able to receive signals from both base stations at the same time. So, we no need to disconnect the old base station before connecting new base station.

2. MOBILITY MANAGEMENT PROTOCOL

Mobility Management is one of the major functions of a GSM or a UMTS network that allows mobile phones to work. The aim of mobility management is to track where the subscribers are, allowing calls, SMS and other mobile phone services to be delivered to them.

Mobile-device technologies have evolved rapidly, allowing users to utilize Internet-based mobility services. IP mobility management protocols are designed to enable the mobile services. The performance of mobility management protocols will have a huge impact on the users' experiences. The performance and features of various mobility-management protocols are important to access in terms of signalling.

2.1 Mobile IPv6

Mobile IPv6 (MIPv6) is a protocol developed as a subset of Internet Protocol version 6 (IPv6) to support mobile connections. MIPv6 is an updated version of the IETF (Internet Engineering Task Force) Mobile IP standard (RFC 2002) designed to authenticate mobile devices (known as mobile nodes) using IPv6 addresses.

In traditional IP routing, IP addresses represent a topology. Routing mechanisms rely on the assumption that each network node will always have the same point of attachment to the Internet, and that each node's IP address identifies the network link where it is connected. In this routing scheme, if we disconnect a mobile device from the Internet and want to reconnect through a different network, we have to configure the device with a new IP address, and the appropriate net mask and default router. Otherwise, routing protocols have no means of delivering data grams (packets), because the device's network address doesn't contain the necessary information about the node's network point of attachment to the Internet.

MIPv6 allows a mobile node to transparently maintain connections while moving from one subnet to another. Each device is identified by its home address although it may be connecting to through another network. When connecting through a foreign network, a mobile device sends its location information to a home agent, which intercepts packets, intended for the device and tunnels them to the current location. Mobility management enables system to locate

roaming terminals in order to deliver data packet and maintain connection with them when moving into new networks.

Handover management is a major component of mobility management since an MN can trigger several handoffs during a session as it is in 4G network. Efficient mechanism mostly ensures seamless handover i.e. with minimal signaling traffic overhead packet delay and lossless. With the connection of various wireless access technologies there are two kinds of handover horizontal handover and vertical handover.

Many of the researchers has adopted for the mobility management like Voice over IP (VOIP) and MIPv6 but there are lot of problems while implementing this mobility management protocol. In this VOIP, it is impossible to talk continuously on mobile device because with mobility the IP changes and hence communication is broken. This problem was solved by using MIPv6 by adopted new mobility management protocol introduced by the researcher. Using MIPv6 enables the MN to communicate with the Corresponding Node (CN) without any break.

When MN is inside its network, it uses home address for communication. When it moves to another network i.e. e.g. GPRS to UMTS it uses Care of Address (CoA). CoA is a temporary address and it is bonded to the MN Home Address (HA).

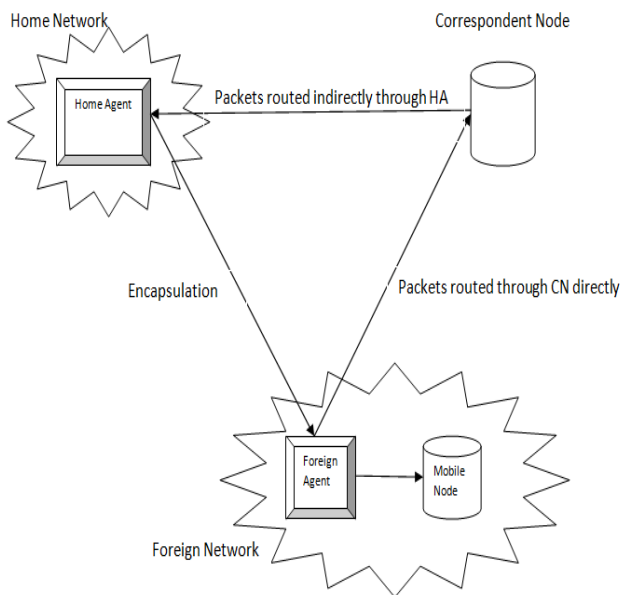


Fig. 3. Triangle Routing

This scheme hides the changed IP from the upper layers. When MN moves from one network to another network CoA is assigned to it by Foreign Agent (FA). When packets

intending the MN arrive HA, it will forward these packets to MN CoA. If MN changes it to CoA, it sends a Binding Update (BU) message to HA and HA reply with BU acknowledgement message. BU updates MN binding information, home address and CoA. When CN sends packet for MN they come to HA sends packet to CN directly with making a triangle routing. Mobile node is able to deliver packets to a CN along a direct path through its FA. Whereas the CA delivers packets to the MN on its home address (means home Network) where Home Agent routes it to the Mobile Node. This asymmetry is known as Triangle Routing, where a single leg of the triangle goes from the MN to the CN, and the HA forms the third vertex controlling the path taken by correspondent node to mobile node. The disadvantage of this routing is that it is not optimal as seen by the long path that a packet will traverse going from sender to MN instead of the direct path from MN to CN.

This way, packets take longer route and network bandwidth can be wasted. To solve this problem MIPv6 is introduced.

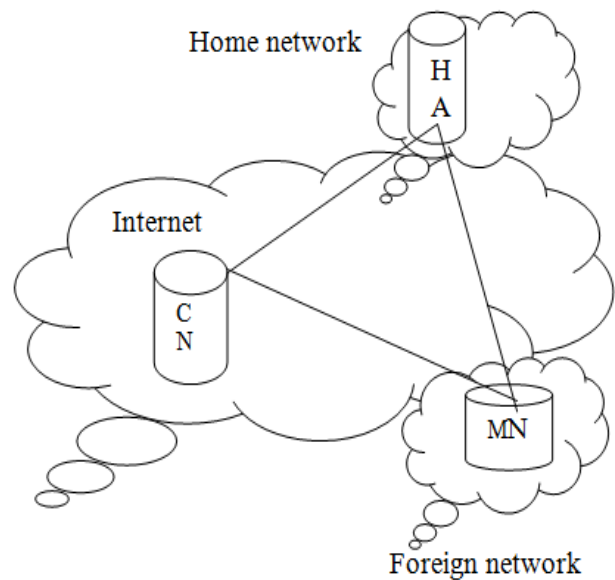


Fig. 4. MIPv6 structure

In this protocol, MIPv6 can keep track of MN CoA by timely BU between MN and its HA, but the problem arises with the packet intended from MN before BU. Discovering a new subnet, establishing a new CoA and information exchange between MN and HA, all the process take time and lot of signalling traffic. Hence it causes high latency and packet loss. The worst case is when MN is roaming between two Access Routers (ARs) several times creating a ping pong effect. In this case too many handover and location updates are experienced and causes interruption in MN communication with it as CN. The packets that were intended for the CoA are dropped. Because of these reason

MIPv6 is not good scheme to perform in 4G high speed data transfer. In order to improve the quality of services of mobility management in 4G network we have to introduce new mobility management to solve this problem.

2.2 Hierarchical Mobile IPv6

Hierarchical Mobile IPv6 (HMIPv6) is the proposed enhancement of Mobile Internet Protocol versions 6 (MIPv6) that is designed to reduce the amount of signalling required and to improve handoff speed for mobile connections. HMIPv6 is a proposed standard from the Internet Engineering Task Force (IETF). MIPv6 defines a means of managing global (between-site) mobility, but doesn't address the issue of local (within-site) mobility separately. Instead, it uses the same mechanisms in both cases, which is an inefficient use of resources in the case of local mobility. HMIPv6 adds another level, built on MIPv6 that separates local from global mobility. In HMIPv6, global mobility is managed by the MIPv6 protocols, while local handoffs are managed locally.

A new node in HMIPv6 called the Mobility Anchor Point (MAP) serves as a local entity to aid in mobile handoffs. The MAP, which replaces MIPv4's foreign agent, can be located anywhere within a hierarchy of routers. In contrast to the foreign agent, there is no requirement for a MAP to reside on each subnet. The MAP helps to decrease handoff-related latency because a local MAP can be updated more quickly than a remote home agent.

Using MIPv6, a mobile node sends location updates to any node it corresponds with each time it changes its location, and at intermittent intervals otherwise. This involves a lot of signalling and processing, and requires a lot of resources. Furthermore, although it is not necessary for external hosts to be updated when a mobile node moves locally, these updates occur for both local and global moves. By separating global and local mobility, HMIPv6 makes it possible to deal with either situation appropriately.

The handoff scheme purposed, HMIPv6 is focused on this paper mainly to solve the problem while MIPv6 is implemented in 4G. The approach of them are in MIPv6 there was no concept of local and global mobility separation, but in the MIPv6 gives the opportunity to deal both this mobility scenarios separately. HMIPv6 fulfils this by introducing a new entity called Mobility Anchor Point (MAP). The global internet is divided into regions; each region is connected to the internet via MAP. It acts as an anchor point to hold the segment together.

In HMIPv6 has use two care of addresses, i.e. Regional Care of Address (RCOA) and the other is Global Care of Address (GCOA). A mobile node communicated with its corresponding node through it RCOA

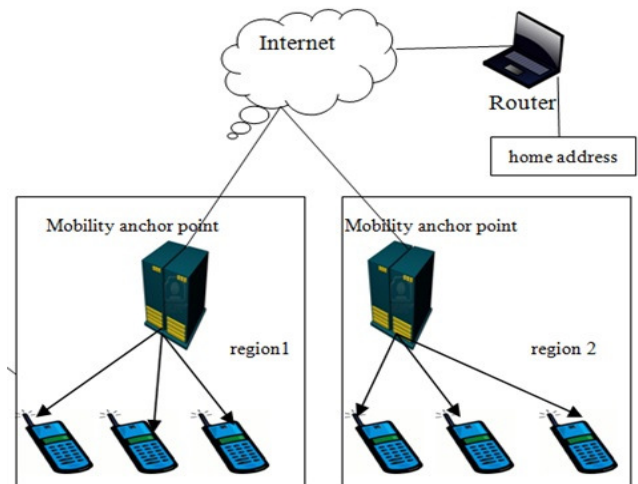


Fig. 5. Simple Structure of HMIPv6

When an MN moves from one network to another network or to a new region, it first takes its RCOA through MAP information. MN then informs it's HA and CN about its point of location. In MIPv6 there was a drawback of repeated connectivity to the same AR. In HMIPv6 when MN repeatedly connected to the same AR Covered by the same MAP, MN takes new COA from MAP called local COA. This mobility is handled locally inside the region and it's reducing the latency factor. As MN COA is changed so the information intended to MN from CN cannot follow the old COA. So HA sends this information to MAP and MAP then tunnels the information to MN local care of address. This way if we implement this type of approached that somewhat increase the quality of service of 4G Networks.

3. RELATED WORKS

As the mobile use increase day by day, the user want not only the communication but also in seamlessly communication during handoff so different researcher adopted research paper for the quality of service improvement in 4G Networks.

In Sarra Benoubira, Mounir Frikha et.al. (2011), they implement Mobile Resource Reservation Protocol (MRSVP) in the fourth generation mobile network to guarantee the QoS for real time application. In this architecture registration of nodes are handled within the hierarchy not need to communicated to the HA. This architecture aim to provide first and efficient handovers between WIMAX and WLAN networks for the mobile node registered in the UMTS network.

Works in Haverinan et.al. (2000) focused on the concept of paging which is one of the main characteristics of Hierarchical Mobile IP. The paging extension that have been used in HMIP allow a mobile node to operate in the power

save mode when locating to another paging area [4]. The location of the mobile node is known as Home agent and represented by paging region. After getting a packet addressed to a mobile node located in a foreign network, the HA tunnels packet to the paging FA, which then the mobile node has been re-established a path towards the current point of attachment.

In Prakash et.al. (2010), they purposed the Hand off Management Unit (HMU) for effective handoff management in 4G Networks. In their research paper they used FBA algorithm which is unbiased and used as a fair approach to resolve the priority issue. The HMU handle effectively horizontal and vertically handoff evaluation through experimentally [1].

In V Savarana's survey paper, he has adopted different types of handoff and handoff factors coming possibility in their survey paper and he also adopted the need of handoff. Again in Mayuri Rao and Manji's survey paper they explained clearly the 4G network technology also in his survey, they mentioned clearly about the 4G wireless systems, its feature and their technology.

Works in Shiva Prasad Kaleru and Damodaram Avula et.al.(2011) introduce the location aided and Route Prediction Methodology of HMIPv6 architecture of 4G Network [10]. They designed a new algorithm Cartesian Co-ordinate system based route prediction algorithm. In this the system is capable of indentifying the geographic location of the network.

Work in Dharminder Kumar and Manoj Yadav clearly mentioned the performance analysis of QoS provided by 3G and 4G Networks in their survey paper [6].

4. PROPOSED SYSTEM

When the mobile node moves from the coverage area of one network to another network, it performs handoff and the connection to the new network domain requires some protocol to minimize the packet loss and latency. We purpose a method for improving QoS during handovers by combining the mobility protocol seamlessly Hierarchy Mobile IPv6 and Session Imitation Protocol (SIP) and then implementing the QoS Manager to the scheme to further increase the security and QoS assurance during handovers.

Our new architecture will contain a SIP server in every domain of the network including the core network. When the MN switches the network it will register to its IP address with the session imitation protocol server available in the new network domain. The MN is communication state with the CN while in handoff. The MN forwards a reconnect message to the SIP proxy server. SIP proxy forwards it to

the CN and they replies an acknowledgement message and call session is established.

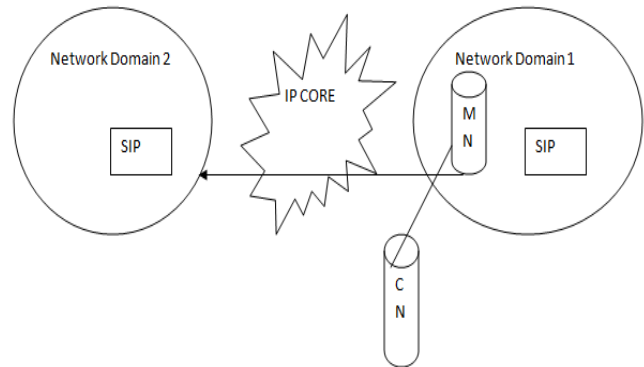


Fig. 6. SIP integrated with HMIPv6

During vertical handoff, the MN needs dynamic allocation and reservation of resources, it is offered by placing QoS manager in QoS framework. Once the handoff is detected by SIP the MN starts sending request to its domain QoS manager. IP core has the capability of allocating resources therefore it is contacted by QoS manger also facilities security by requesting Authentication, Authorization and Accounting server (AAA). If QoS manager feel that during resource allocation there is deficiency it will contact with neighbour QoS manager for help. In this way SIP Proxy provide in minimizing packet losses and jitters.

5. CONCLUSION

Through this propose of this paper the HMIPv6 for effective handoff management in 4G networks. In the 4G Networks would revolutionize the future networks providing better QoS, superior integrated service and mobility to the user. A lot of research is going on to realize 4G networks to improve the quality of service using mobility management protocol.

Mobile communication has become the important aspect of the day today lives and its importance and scope cannot be denied in our lives. Mobile users demand for advance techniques with efficiency and sophistication. Fourth generation of mobile networks is developed to meet the requirements of users in data rates and speed. While keeping QoS assurance in focus these requirements are needed to be fulfilled.

The main problems that 4G is facing are seamless communication, security and QoS assurance. These areas need to be addressed and improved if 4G wants to become the most advance technology in telecommunications.

We present the challenges that 4G faces and their up-to-date solutions. To improve the QoS in 4G we propose our own

scheme of combining mobility protocol HMIPv6 and application layer protocol SIP. With this scheme, the QoS level in 4G can be improved because both the protocols provide support in handovers. Together they can decrease the packet loss and can improve security during the handover process. We can make sure the resource allocation during the handover process by combining the two protocols and mobility management can be optimized.

6. ACKNOWLEDGMENT

The author would like to express their cordial thanks to Miss Himanshu Sharma – Lovely Faculty of Tech. And Science and my friends Barun Khomdram, Robert Hijam, Rajshree Rajkumari, Ibehaibi Sagolsem for their valuable advice and expertise in the area of 4G Networks.

REFERENCES

- [1] Parkash.s, C.B.Akki, Kashyap Dhruve handoff management Architecture for 4g network over MIPv6, ICSNS, and VOL.10 No 2. Feb 2010
- [2] Christian and sammuel, Handoff Protocol for heterogeneous all IP base wireless network,
- [3] Xiaoming fu, Rene Soltwisch, Qos and security in 4g networks
- [4] DR. Manjaiah, D.H Payaswini, Challenge and issue in Mobile IPv6 based mobility management Approaches of 4G Network.
- [5] Aisha H.A Hasim,Hatina Liyakthalik, Mobility issue in hierarchy Mobile IP,SETIT, March 27-31,2005-TUNISIA
- [6] Dharamander Kumar and Manoj Yadav, performance Analysis of Quality of services provided by 3G and 4G Network, 5 National conference ;INDIA Com -2011
- [7] Mayuri Rao, Manish Panchel. 4G Wireless Technologies, NCNTE -2012, Vashi, Navi Mumbai, Feb 24-25, 2012
- [8] Firas Qusta, Nidal kamel, Mohd Zuki, Charles Asrraf. Optimization of Quality of Service in 4G wireless Networks, ACEEE on network security, vol 03, No. 02, April 2012
- [9] Sarra Benoubira, Mounir Frikha, Sami Tabbane, Mobility and QoS Management in Heterogeneous Wireless Networks, IJCSNS, VOL 11 No 9, September 2011
- [10] Shiva Prasad kaleru and Damodaram Avula, Location Aided HMIPv6 Architecture for Vertical Handoff in 4G Network, Euro Journals Publishing inc 2011.

Use of Storage as Service for Online Operating System in Cloud Computing

Piyush Saxena¹, Satyajit Padhy², Praveen Kumar³

¹M.Tech (CS&E), Amity University, Noida, India, piyushisgenius@gmail.com

²M.Tech (CS&E), Amity University, Noida, India, satyajitpadhyewit@gmail.com

³Assistant Professor, Amity University, Noida, India, pkumar3@amity.edu

Abstract: Cloud computing has made it possible to make system boot using online operating system thus saving both primary and secondary memory because the data is on centralized data centre located outside the organization which is highly secure. It is not in computer memory so that it can be accessed anywhere. It also saves money as one doesn't need to buy any expensive hardware to access the particular software in your computer. Cloud computing is a highly scalable pay-per-use IT capabilities. Now a days, software is very much expensive which even MNC's don't want to purchase it due to various factors which is:-Not reliable, Highly expensive which is very costly to install it on 1000's of computers, If any error occurrence it takes 1-2 days to solve which is a big loss for organization. So, here is the simple solution i.e. cloud computing which makes organization more productive due to low cost of software with high-end features, highly reliable, low maintenance cost, problems solving immediately and Highly secure.

Cloud computing allows consumers and businesses to use application without installation and access their personal files at any computer with internet access. Speed up the calculations and processes. Provide multi tenancy features. Cloud Computing is a Computing in which services and software are provided over the Internet ("cloud") which is very cheap and affordable. Cloud computing is on demand access to virtualized IT resources that are housed outside of your own data centre, shared by other simple to use, paid for via monthly subscription which is very low in cost, and accessed over the web with many features in it. Storage as a Service is a business model in which a large company rents space in their storage infrastructure to another company or individual. In the enterprise, StaaS vendors are targeting secondary storage applications by promoting StaaS as a convenient way to manage backups. The key advantage to StaaS in the enterprise is in cost savings -- in personnel, in hardware and in physical storage space. The StaaS provider agrees to rent storage space on a cost-per-gigabyte-stored and cost-per-data-transfer basis and the company's data would be automatically transferred at the specified time over the storage provider's proprietary wide area network (WAN) or the Internet. If the company's data ever became corrupt or got lost, the network administrator could contact the StaaS provider and request a copy of the data.

Index Terms: Virtualized IT Resources, StaaS, cloud computing

1. INTRODUCTION

Cloud computing is known as Internet based computing, with shared resources, software and information is provided to computers and other devices. Cloud Computing is a Computing in which services and software are provided over the Internet ("cloud") which is very cheap and affordable.

Cloud computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. Example of cloud computing is Yahoo mail or Gmail etc. You don't need software or a server to use them. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider Yahoo, Google etc. The consumer gets to use the software alone and enjoy the benefits. Cloud computing is on demand access to virtualized IT resources that are housed outside of your own data centre, shared by others, simple to use, paid for via monthly subscription which is very low in cost, and accessed over the web with many features in it.

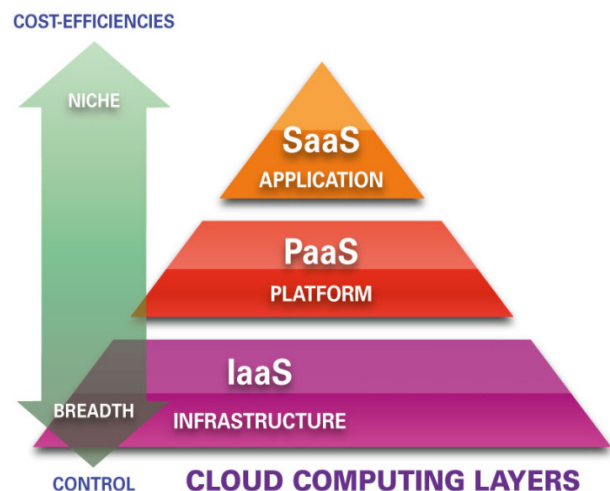


Fig. 1. Types of Cloud Services

Cloud computing makes use of SaaS. Storage as a Service is generally seen as a good alternative for a small or mid-sized business that lacks the capital budget and/or technical personnel to implement and maintain their own storage infrastructure. SaaS is also being promoted as a way for all businesses to mitigate risks in disaster recovery, provide long-term retention for records and enhance both business continuity and availability. Cloud storage is a service model in which data is maintained, managed and backed up remotely and made available to users over a network (typically the Internet).

There are three main cloud storage models:[4]

1. Public cloud storage services, that provides a multi-tenant storage environment that's most suitable for unstructured data.
2. Private cloud storage services provide a dedicated environment protected behind an organization's firewall. Appropriate for users who need customization and more control over their data.
3. Hybrid cloud storage is a combination of the other two models that includes at least one private cloud and one public cloud infrastructure.

By making data available in the cloud, it can be more easily and ubiquitously accessed, often at much lower cost, increasing its value by enabling opportunities for enhanced collaboration, integration, and analysis on a shared common platform.

The basic concept of cloud computing is using software via the Internet instead of installing it onto personal computer. From this cloud, one can find programs that he/she wants to utilize. Through web browser anyone can access programs from cloud.

2. CLOUD SERVICES

Business applications like those from SAP, Microsoft, and Oracle have always been too complicated and expensive. They need a data centre with office space, power, cooling, bandwidth, networks, servers, and storage etc. And a team of experts to install, configure, and run them. And a complicated software stack. It multiplies the headaches when this takes place across dozens or hundreds of apps, it's easy to see why the biggest companies with the best IT departments isn't getting the applications they need.

By cloud computing and IT as a Service (ITaaS) could bring the total costs down. It would reduce the need for advanced hardware on the client side. One wouldn't need to buy the fastest computer with the most memory, because the cloud system would take care of those needs.

IT as a service can be of 4 types:-

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)
4. Storage as a Service (StaaS)

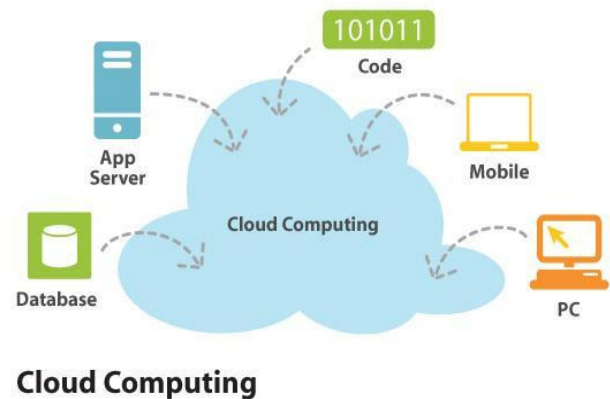


Fig 2. Implementation in Cloud

Description of the Services in Detail:

1) **Infrastructure as a Service:** IaaS manages a large set of computing resources, such as sorting and processing capacity. Through virtualization, they are able to split, assign and dynamically resize these resources to build ad-hoc systems as demanded by customers. They deploy the software stacks that run their services. IaaS is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. Example: go grid etc.

2) **Platform as a Service:** Cloud systems can offer an additional abstraction level: instead of supplying a virtualized infrastructure, they can provide the software platform where systems run on. The sizing of the hardware resources demanded by the execution of the services is made in a transparent manner. This is denoted as Platform as a Service (PaaS). Example: Google Apps Engine etc.

3) **Storage as a Service:** [3] Commonly known as Storage as a Service (StaaS), it facilitates cloud applications to scale beyond their limited servers. StaaS allows users to store their data at remote disks and access them anytime from any place. Cloud storage systems are expected to meet several rigorous requirements for maintaining users' data and information, including high availability, reliability, performance, replication and data consistency; but because of the conflicting nature of these requirements, no one system implements all of them together.

4) *Software as a Service*: Finally, there are services of potential interest to a wide variety of users hosted in Cloud systems. SaaS sometimes referred to as "software on demand". It is software that is deployed to run over the internet or behind a firewall on a local area network or personal computer. SaaS has become a common model for many business applications including accounting, collaboration, customer relationship management (CRM), enterprise resource planning (ERP), invoicing, human resource management (HRM), content management (CM) and service desk management. Example: Salesforce.com etc.



Fig. 4. Overview of a cloud

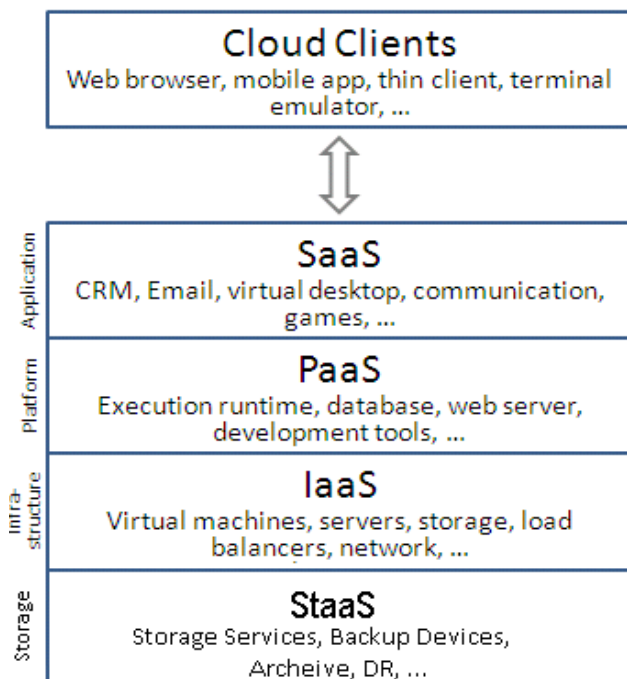


Fig. 3. Phases of services in cloud

Some Characteristics of the Cloud System are:

1. **Data Durability and Reliability** - It stands that one can fully trust and rely on these kinds of services.
2. **Security** - "due to centralization of data"...There is very less chance to hack and occurring some problems. In case there is any problem it can be resolved immediately.
3. **Privacy** -There is no interconnection between two clients. It provides username and password which kept safe and the clients can change the password anytime. (If they doubt that their password has been breached).
4. **Reduced Redundancy storage** - Reduces the redundant data stored and thus gives an excellent utilization of storage.
5. **Backup, Archiving and Disaster Recovery** - Backups the complete data of data centres on many servers so as to keep up a disaster recovery logs and data for in case something happens. This is ideal for moving large quantities of data for periodic backups, or quickly retrieving data for disaster recovery scenarios.

Some Advantages of using Cloud Computing Systems are:-

1. **Low use of power supply** as the power used to run multiple apps now is used for a single app which reduces power supply.
2. **Reduces the need of high levels of cooling** as only one app needs to run as compared to multiple apps.
3. **Bandwidth problem**-it relates to networks in which there is no congestion or traffic in network signal basically due to centralized server.

3. MULTI-TENANCY

Every corporation shares same application which they can customize with their own specific needs. As the data of every corporation is private it can't be access by any other corporation because it is highly secure services which is password protected. Only administrator can access the data into the server held in cloud.

"Multi-Tenancy" indicates that some infrastructure is shared, at what layers are things being shared can make a big difference. For example, Amazon AWS is multi-tenant at the hardware level in that its users may be sharing a physical machine. On the other hand, Force.com is multi-tenant at the DB level in that its users are sharing data in the same DB tables. And Amazon is relying on the hypervisor to provide

the isolation between tenants while Force.com is relying on a query rewriter to do the same.

4. CLOUD STORAGE

Cloud storage is amorphous today, with neither a clearly defined set of capabilities nor any single architecture. Choices abound, with many traditional hosted or managed service providers (MSP) offering block or file storage, usually alongside traditional remote access protocols or virtual or physical server hosting. Cloud storage can be defined as a specific category within the larger field of — storage in the cloud solutions.

Storage in the cloud encompasses traditional hosted storage, including offerings accessed by FTP, WebDAV, NFS/CIFS, or block protocols either remotely [3] or from within a hosted environment. Cloud storage is an evolution of this hosted storage technology that wraps more sophisticated APIs, namespaces, file or data location virtualization, and management tools, around storage.

The evolution of Cloud Storage is based on traditional network storage and hosted storage. There are hundreds of different cloud storage systems. Some have a very specific focus, such as storing Web email messages or digital pictures. Others are available to store all forms of digital data. Some cloud storage systems are small operations, while others are so large that the physical equipment can fill up an entire warehouse. The facilities that house cloud storage systems are called data centres.

At its most basic level, a cloud storage system needs just one data server connected to the Internet. A client sends copies of files over the Internet to the data server, which then records the information. When the client wishes to retrieve the information, he or she accesses the data server through a Web-based interface. The server then either sends the files back to the client or allows the client to access and manipulate the files on the server itself.

5. RISKS RELATED TO CLOUD COMPUTING

Main fact that cloud computing services are shared all over the world and could be shut down at any time. If any problem in cloud computing server it will affect complete clients connected to that server. The basic fact that everyone is sharing data with world in big security risk.

6. CLOUD BASED OPERATING SYSTEM

A Web Operating System is a Web platform which allows the user to use a virtual Desktop through a web browser rather than using any particular local operating system. This amazing technology allows a user to access their own virtual

desktop from anywhere around the world, without even using a network like with a remote PC.

In addition, you are essentially using the Internet to work on a virtual desktop, rather than working on an actual desktop computer.

Some examples of cloud based operating system are:

1. <http://www.Beta.cloudo.com>
2. <http://www.Oos.cc>
3. <http://www.eyeos.info>
4. <http://www.lucid-desktop.org/>
5. <http://www.amoebaos.org/>

7. CANDIDATES OF CLOUD IMPLEMENTATION

One of the reasons the idea of cloud computing is getting so much traction is the efforts of companies other than Google to make it happen.

Some of those known to be considering providing Cloud Services are:

- Microsoft
- Amazon
- IBM
- Google
- Apple

In its Blue Cloud initiative, IBM wants to network its massive computers, like this one at a research centre in Jülich, Germany, to create a powerful cloud.



Fig. 5. Organisations using cloud

8. IMPLEMENTATION

Among the most popular cloud services now are social networking sites (the 500 million people using *Facebook* are being social in the cloud), webmail services like Hotmail and Yahoo mail, micro blogging and blogging services such as Twitter and Word Press, video-sharing sites like YouTube, picture-sharing sites such as Flickr, document and applications sites like Google Docs, social-bookmarking sites like Delicious, business sites like eBay, and ranking, rating and commenting sites such as Yelp and Trip Advisor.

9. FUTURE ASPECTS

"By 2020, most people won't do their work with software running on a general-purpose PC. Instead, they will work on Internet-based applications such as Google Docs, and in applications run from smart phones. Aspiring application developers will develop for Smartphone vendors and companies that provide Internet-based applications, because most innovative work will be done in that domain, instead of designing applications that run on a PC operating system."

Most of those surveyed noted that cloud computing will continue to expand and come to dominate information transactions because it offers many advantages, allowing users to have easy, instant, and individualized access to tools and information they need wherever they are, locatable from any networked device. Some experts noted that people in technology rich environment will have access to sophisticated-yet-affordable local networks.

As new offerings like Amazon's Cloud Front, Microsoft's Azure, Hosting.com's Cloud Nine and VMware's vCloud are rolled out week in, week out; the worldwide cloud computing momentum continues to grow. Here, SYS-CON's Cloud Computing Journal surveys a globe – girdling network of leading infrastructure experts, IT industry executives and technology commentators for their on The Shape of Cloud Computing to Come.

Contributors include:-

1. Salesforce.com's Peter Coffee
2. Geve Perry of GigaSpaces
3. Ben Rushlo from Keynote Systems
4. Cloud Computing Journal editor-in-chief Alan Williamson
5. Founder Reuven Cohen, open source entrepreneur
6. Krishnan Subramanian and Markus Klems of the FZI Research Centre or Information Technology in Germany.

10. CONCLUSIONS

Cloud Storage with a great deal of promise, aren't designed to be high performing file systems but rather extremely scalable, easy to manage storage systems. They use a different approach to data create a very scalable storage system. Typically cloud computing provides cost effective redundancies in storage hardware. This translates into uninterrupted service during a planned or unplanned outage. This is also true for hardware upgrades which for the end user will no longer be visible resiliency, Redundant array of inexpensive nodes, coupled with object based or object-like file systems and data replication (multiple copies of the data), to form a cloud system.

REFERENCES

- [1] Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide by David S. Linthicum.
- [2] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing," in Proc. of IWQoS'09, July 2009, pp.1–9.
- [3] A. Juels and J. Burton S. Kaliski, "Pors: Proofs of retrievability for large files," in Proc. of CCS'07, Alexandria, VA, October 2007, pp. 584–597.
- [4] Jiyi Wu, Lingdi Ping, Xiaoping, Ya Wang, Jianqing, 2010 International Conference on Intelligent Computing and Cognitive Informatics, — Cloud Storage as the Infrastructure of Cloud Computing.
- [5] K. D. Bowers, A. Juels, and A. Oprea, "Hail: A high-availability and integrity layer for cloud storage," in Proc. of CCS'09, 2009, pp. 187–198.
- [6] <http://searchsmbstorage.techtarget.com/feature/Understanding-cloud-storage-services-A-guide-for-beginners>
- [7] Sun Microsystems, Inc., "Building customer trust in cloud computing with transparent security," Online at <https://www.sun.com/offers/details/sun-transparency.xml>, November 2009.
- [8] Storage Networking Industry Association. Cloud Storage Reference Model, Jun.2009.
- [9] <http://www.business.att.com/enterprise/Service/hosting-services/cloud/storage/>
- [10] Storage Networking Industry Association. Cloud Storage for Cloud Computing, Jun.2009
- [11] Cloud computing: state-of-the-art and research challenges By- Qi Zhang, Lu Cheng, Raouf Boutaba Journal of Internet Services and Applications May 2010, Volume 1, Issue 1, pp 7-18
- [12] L. Wang, J. Zhan, W. Shi, Y. Liang, and L. Yuan, "In cloud, do mtc or htc service providers benefit from the economies of scale?" in SC-MTAGS, 2009.
- [13] A view of cloud computing By- Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, Matei Zaharia Published in: Communications of the ACM CACM Homepage archive Volume 53 Issue 4, April 2010 Pages 50-58

- [14] Cloud Computing: Distributed Internet Computing for IT and Scientific Research Internet Computing, IEEE Date of Publication: Sept.-Oct. 2009 Author(s): Dikaiakos, M.D. Univ. of Cyprus, Nicosia, Cyprus Katsaros, D. ; Mehra, P. ; Pallis, G.; Vakali, A. Volume: 13 , Issue: 5
- [15] The Case for Cloud Computing IT Professional Date of Publication: March-April 2009 Author(s): Grossman, R.L. Volume: 11, Issue: 2 Page(s): 23 – 27
- [16] Service-Oriented Computing and Cloud Computing Challenges and Opportunities Yi Wei and M. Brian Blake University of Notre Dame.
- [17] Cloud Computing Authors: Greg Boss, Padma Malladi, Dennis Quan, Linda Legregni, Harold Hall, Date: 8 October 2007 Status: Version 1.0 Copyright IBM Corporation 2007
- [18] What's inside the Cloud? An architectural map of the Cloud landscape By-Alexander Lenk, Markus Klems, Jens Nimis, Stefan Tai, Thomas Sandholm Published in: CLOUD '09 Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing Pages 23-31 IEEE Computer Society Washington, DC, USA ©2009

Overview of Motion Estimation Algorithms in Video Compression for H.261 ITU-I Recommendation

Rahul Bhandari¹, Eshank Jain²

^{1,2}M. Tech. Student, CSE Department, JIET, Jodhpur, Rajasthan, India
¹rb_rahulbhandari@rediffmail.com, ²eshank.jain@jiethodhpur.com

Abstract: Motion estimation is one of the most important techniques in video compression. It can define as encoder searching a previous block of pixels to determine where a block of pixels in the current frame has moved. If successfully, encoder via motion compensation can then transmit only difference information between two frames across the transmission channel. This saves a significant amount of channel bandwidth. The problem with motion estimation lies in the size of the problem. Even modest areas to search increase in computational complexity rapidly, which leads to large increases in power consumption. A variety of algorithms and techniques have been developed to find the optimal motion vector given the large throughput demands of the operations.

1. INTRODUCTION

Compression is the process of compacting data into a smaller number of bits. Video compression (video coding) is the process of compacting or condensing a digital video sequence into a smaller number of bits. 'Raw' or uncompressed digital video typically requires a large bit rate (approximately 216 Mbits for 1 second of uncompressed TV-quality video) and compression is necessary for practical storage and transmission of digital video. Compression involves a complementary pair of systems, a compressor (encoder) and a de-compressor (decoder). The encoder converts the source data into a compressed form (occupying a reduced number of bits) prior to transmission or storage and the decoder converts the compressed form back into a representation of the original video data. The encoder/decoder pair is often described as a *CODEC* (enCOder/ DECOder) [10].

Data compression is achieved by removing redundancy, i.e. components that are not necessary for faithful reproduction of the data. Many types of data contain statistical redundancy and can be effectively compressed using lossless compression, so that the reconstructed data at the output of the decoder is a perfect copy of the original data. Unfortunately, lossless compression of image and video information gives only a moderate amount of compression. Lossy compression is necessary to achieve higher compression. In a lossy compression system, the

decompressed data is not identical to the source data and much higher compression ratios can be achieved at the expense of a loss of visual quality. Lossy video compression systems are based on the principle of removing subjective redundancy, elements of the image or video sequence that can be removed without significantly affecting the viewer's perception of visual quality [10].

Most video coding methods exploit both temporal and spatial redundancy to achieve compression. In the temporal domain, there is usually a high correlation (similarity) between frames of video that were captured at around the same time. Temporally adjacent frames (successive frames in time order) are often highly correlated, especially if the temporal sampling rate (the frame rate) is high. In the spatial domain, there is usually a high correlation between pixels (samples) that are close to each other, i.e. the values of neighbouring samples are often very similar.

The basic objective of compression of any sort is to reduce the required bit rate by removing redundant information carried in a particular form. For video images, like other compressed data, there are mainly two sorts of video compression/coding schemes:

2. SOURCE CODING & ENTROPY CODING

Source coding deals directly with characteristics of the source material and yields results which are lossy, that is, picture quality is degraded, while entropy coding relies on the statistical properties of the signals and, is lossless [1]. H.261 makes use of both techniques.

It can be divided into two further types: intra-frame and inter-frame. Intra-frame coding is used for the first picture in a sequence and for images after a scene changed, inter-frame coding is used for sequence of similar images, including those containing moving objects. Intra-frame coding removes redundant spatial information, also removes temporal redundancy between images.

H.261 makes use of both source and entropy coding as well as inter and intra variations of source coding. The data rate of the coding algorithm was designed to be able to be set to between 40 Kbits/s and 2 Mbits/s. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codec compensate for motion. To remove any further redundancy in the transmitted bit stream, variable length coding is used.

H.261 supports two resolutions, QCIF (Quarter Common Interchange format) and CIF (Common Interchange format) and The Image quality measure used is the peak signal to noise ratio (PSNR).

Table I: Picture Formats Supported [1]

Picture format	Luminance pixels	Luminance lines	H.261 support	Uncompressed bit rate (Mbit/s)			
				10 frames/s		30 frames/s	
				Grey	Colour	Grey	Colour
QCIF	176	144	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	8.1	12.2	24.3	36.5

A higher PSNR would normally indicate that the reconstruction is of higher quality. Acceptable values for wireless transmission quality loss are considered to be about 20 dB to 25 dB.

MSE=

$$\frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N [I(x,y) - I'(x,y)]^2$$

$$\text{PSNR} = 20 * \log_{10} (255 / \sqrt{\text{MSE}})$$

where $I(x, y)$ is the original image, $I'(x, y)$ is the approximated version (which is actually the decompressed image) and M, N are the dimensions of the images. A lower value for MSE means lesser error, and as seen from the inverse relation between the MSE and PSNR, this translates to a high value of PSNR. Logically, a higher value of PSNR is good because it means that the ratio of Signal to Noise is higher. Here, the 'signal' is the original image, and the 'noise' is the error in reconstruction. So, if you find a compression scheme having a lower MSE (and a high PSNR), you can recognise that it is a better one [1].

Entropy coding is a loss-less compression scheme based on statistical properties of the picture or the stream of the information to be compressed [1]. Although entropy coding is implemented slightly different in each of the standards,

the basic “entropy coding” scheme consists of encoding the most frequently occurring patterns with the least number of bits. Most commonly data can compress by an additional factor of 3 or 4.

Entropy coding for video compression applications is a two step process: Zero Run-Length Coding (RLC) and Huffman coding.

RLC data is an intermediate symbolic representation of the quantized bins which utilizes a pair of numbers. The first number represents the number of consecutive zeros while the second number represents the value between zero-run lengths. For instance the RLC code (5, 8) represents the sequence (0, 0, 0, 0, 0, 8) of numbers.

Huffman coding assigns a variable length code to the RLC data, producing variable length bit stream data. This requires Huffman tables which can be pre-computed based on statistical properties of the image (as it is in JPEG) or can be pre-determined if a default table is to be used (as it is in H.261 and MPEG). In either case, the same table is used to decode the bit stream data.

3. INTRODUCTION TO H.261 ENCODER

H.261 is the ITU-T telecommunication standardization which is used to video telephony, video conferencing etc.

Several characteristics of H.261 are: (1) H.261 comprises between coding performance, real time requirements, implementation complexities, and system robustness. (2) it coding structures and parameters are tuned towards low-bit rate transmission. (3) it designed for real time communications and to reduce encoding delay uses closest previous frame for motion picture sequence coding.

H.261 specifies a set of protocols that every compressed-video bit-stream must follow and a set of operations that every standard. The data structure of the encoder / decoder and the requirements of the video bit-stream also are described. The video bit-stream contains the picture layer, group-of-blocks layer, macro-block layer, and the block layer [11].

Picture layer: Data for each picture consists of a picture header followed by data for a GOB.

GOB layer: Each picture is divided into GOBs, each of which is one-twelfth of the CIF or one-third of the QCIF picture area. A GOB relates to 176 pixels by 48 lines of Y and the spatially corresponding 88 pixels by 24 lines for each C_b and C_r .

Macro-block (MB) layer: Each GOB layer is divided into 33 macro-blocks. A macro-block relates to 16 pixels by 16

lines of Y and the spatially corresponding 8 pixels by 8 lines for each C_r and C_b .

Block layer: A macro-block is composed of four luminance blocks and one each of the two color difference blocks. Data for a block consists of code-words for transform coefficients followed by an end-of-block (EOB) marker [11].

4. WORKING OF H.261

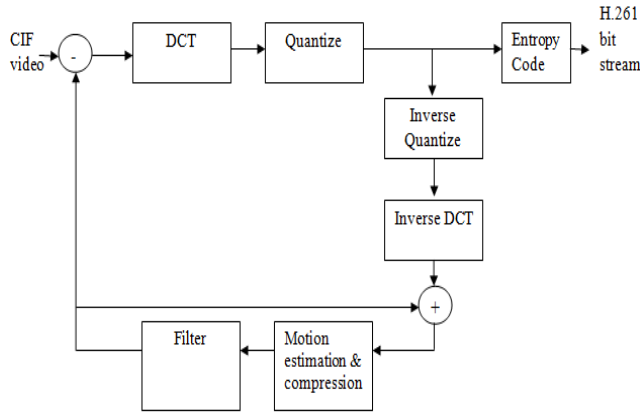


Fig. 1. H.261 Encoder [11]

First, motion estimation is performed on each macro block. Since objects in the frame may be moving in different directions, each macro block is allowed to have a different motion vector. The motion vector is used as a displacement vector to fetch a macro block from the preceding frame to be used as a prediction.

Motion estimation in H.261 is only performed relative to the preceding frame, and on full-pixel offsets up to a maximum of ± 15 in the horizontal and vertical directions. To improve the prediction, H.261 allows for an optional loop-filter to be applied to the prediction on a macro block basis [11].

Next, a decision must be made to code either the arithmetic difference between the offset prediction macro block and the current macro block or to code the current macro block from scratch. Since the arithmetic difference is usually small, coding the arithmetic difference results in higher compression [11].

An 8x8 DCT is applied to each block in either the arithmetic difference macro block or the current macro block. Instead of quantization matrices, H.261 uses one quantization scale for all frequency bins.

Since the DC bin is the most important, it is separately quantized to a fixed 8 bit scale. Adjustment of the quantization scale on a per macro block basis is the primary

method for controlling the quality and compression ratio in H.261 [11].

5. MOTION ESTIMATION

In general, successive pictures in a motion video sequence tend to be highly correlated, that is, the pictures change slightly over a small period of time. This implies that the arithmetic difference between these pictures is small. For this reason, compression ratios for motion video sequences may be increased by encoding the arithmetic difference between two or more successive frames.

In contrast, objects that are in motion increase the arithmetic difference between frames which in turn implies that more bits are required to encode the sequence. To address this issue, motion estimation is utilized to determine the displacement of an object.

Motion estimation is the process by which elements in a picture are best correlated to elements in other pictures (ahead or behind) by the estimated amount of motion. The amount of motion is encapsulated in the motion vector. Forward motion vectors refer to correlation with previous pictures. Backward motion vectors refer to correlation with future pictures [8].

An efficient motion estimation algorithm increases frame correlation, which in turn minimizes pixel arithmetic difference. Resulting in not only higher compression ratios but also in higher quality decoded video sequences. Motion estimation is an extremely computationally intensive operation difficult to implement in real-time. For this reason, a variety of motion estimation algorithms have been implemented by the industry [8].

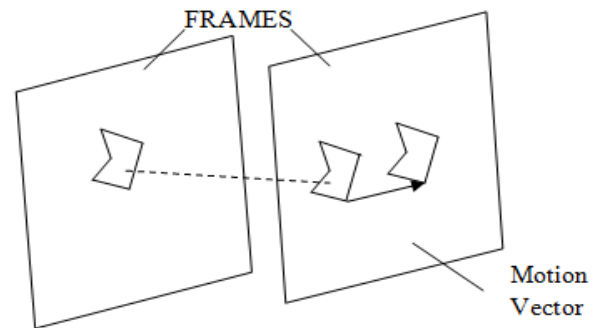


Fig. 2. Motion Estimation [8]

There are two fundamentally different types of motion estimation algorithms: block-matching algorithms in the spatial domain and correlation algorithms in the two-dimensional (spatial) frequency domain [8]. Block-matching algorithms are mathematically simpler, but require high processing power to achieve a large search range.

Correlation algorithms can achieve higher (sub-pixel) accuracy and can provide multiple motion vector candidates in one calculation [8].

6. BLOCK-MATCHING ALGORITHMS IN THE SPATIAL DOMAIN

1. Exhaustive motion estimation:

The simplest, and also the most processing-intensive, block-matching algorithm is one that is exhaustive or enables full search over the entire search range [9].

Figure 3 shows the frame order and the motion estimation from a reference frame to a coded P frame.

Figure 4 shows exhaustive search algorithm. Using the sum of absolute differences (SAD) between the predicted block and the search block as a match criterion, exhaustive motion estimation examines all possible search positions within a pre-defined search area and uses the one with the lowest SAD (best match) as the prediction candidate.

Similarly, Figure 5 shows the motion estimation process on a B frame. Since the processing power requirements of exhaustive motion estimation increases with the square of the search range, the question arises as to how big the search range needs to be. Examining fast sports material on a frame-by-frame basis, it can easily be seen that motion displacements of 100 pixels per frame.

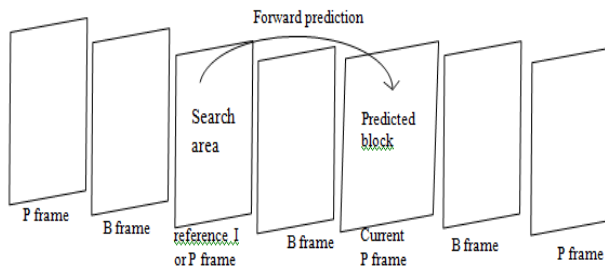


Fig. 3. motion estimation from reference frame to current P frame

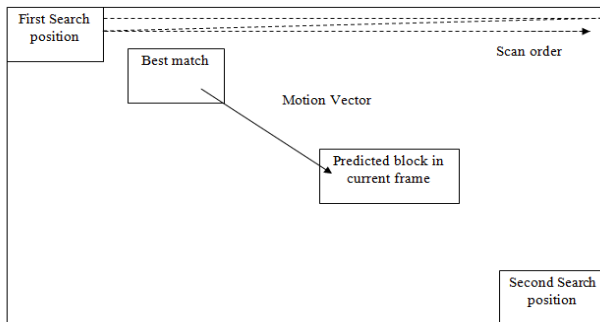


Fig. 4. Exhaustive search algorithm

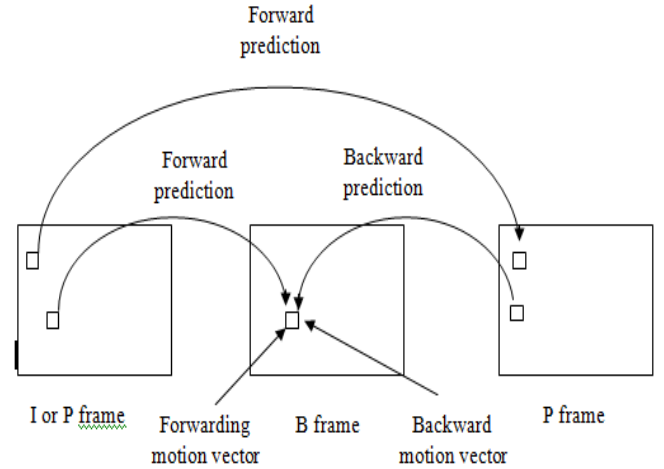


Fig. 5. Motion estimation on B-frame

2. Hierarchical motion estimation:

One of the most popular methods for reducing the processing requirements of block-matching algorithms is to downsize the image several times to produce a hierarchical structure of images. Down-sampling reduces computational requirements for two reasons. Not only are the block sizes smaller, cutting down on the number of calculations necessary for each search position, but pixel displacements in the smaller images also correspond to larger displacements in the base layer and, therefore, a small search range in the smallest image corresponds to a larger search range in the base layer.

Fig. 6 shows a block diagram of a hierarchical motion estimation system. The 1:2 down-sample filter blocks are re-sizing both the reference image and the search image in both horizontal and vertical directions. Each down-sample block therefore represents four one-dimensional down-sampling filters. The motion estimation process starts off at the smallest image to produce candidate vectors MV1 for next larger image. At each stage, the vectors found at the higher layer are used as a starting point for a refinement search at the current layer, whereby the refinement search range is much smaller than the total search range, typically only a small number of pixel positions in each direction. At the bottom layer motion vectors MV4 are the pixel-accurate vectors used for motion compensation.

While hierarchical motion estimation can give good performance in many applications where processing power is at a premium, there are a number of disadvantages and potential problems with this form of motion estimation. The most obvious deficiency arises from the fact that the search in the higher layers can get trapped in local minima and it is often difficult to recover from sub-optimum starting points at the lower layers.

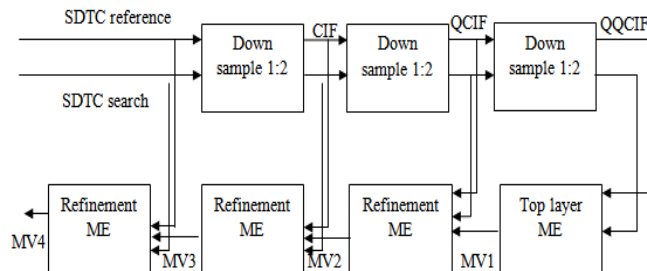


Fig. 6. Block diagram of a hierarchical motion estimation system

A second disadvantage in a real-time implementation of hierarchical motion estimation is the fact that the higher layers of the motion estimation process have to be carried out on the source picture because in a real-time system there is not enough time to do a search on all layers of the reconstructed picture, i.e. within the coding loop.

7. CONCLUSION

This paper describes the overview of different algorithms of Block Matching Algorithms for compressing a raw video. This paper also shows that while compressing a video, motion estimation has very vital role. When motion estimation is compressed then real time compression of video can achieved.

REFERENCES

- [1] Video Codec for Audiovisual Services at px64 kbits/sec – Recommendation H.261.
- [2] Hung, Andy C., “PVRG-p64 Codec 1.1 Manual, Nov17, 1993 (<http://havefun.stanford.edu>).
- [3] Texas Instruments On-line application notes.
- [4] Compression Technology: an MPEG overview” – C-Cube Microsystems Application notes (<http://www.ccube.com>).
- [5] Source Code courtesy from (<ftp://havefun.stanford.edu>).
- [6] Data and Image Compression: Tools and Techniques; Held, Gilbert/ Marshall, Thomas.
- [7] H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia: Iain E. G. Richardson, The Robert Gordon University, Aberdeen, UK.
- [8] A New Motion Compensation Algorithm in DCT Domain for H.261 Video Encoder*, Xiangyang Xue and Changxin Fan7 FIEEE, Information Science Institute, Xidian University, Wan, 710071, China.
- [9] “A Low-Power Design of Motion Estimation Blocks for Low Bit-Rate Wireless Video Communications”, Steve Richmond, Dr. Dong S. Ha, Chairman, Bradley Department of Electrical and Computer Engineering, March 2001 Blacksburg, Virginia.
- [10] “Video compression system for first principles to concatenated codecs”, Alios M. Bock, Telecommunication Series 53.
- [11] “Video codec for audiovisual services at p x64 kbit/s”, ITU-T Recommendation H.261
- [12] “H.261 Implementation on the TMS320C80 DSP”, Texas Instruments, SPRA161, June 1997.

Investigation of XGM effect of SOA for OOK, DPSK, Duo-binary & Manchester format

Manish Chauhan¹, Hemant Purohit²

¹M.Tech. Student, ECE Department, JIET, Jodhpur, Rajasthan, India

²Associate professor, EEE Department, JIET-SETG, Jodhpur, Rajasthan, India

¹Manish.chauhan@jiетjodhpur.com, ²hemant.purohit@jiетjodhpur.com

Abstract: Due to different modulation formats, Conventional ON-OFF keyed (OOK) signals such as return-to-zero (RZ) and non return-to-zero (NRZ) signals are susceptible to cross-gain modulation (XGM) in semiconductor optical amplifiers (SOAs) that leads to crosstalk penalty for multiplexed input signals [1]. Differential phase-shift keying of optical pulses has virtually no pulse-pattern effect and are robust to cross-gain modulation [2]. Here we will compare Cross Gain Modulation (XGM) effect of OOK, Manchester, duo-binary & differential phase-shift keying (DPSK) signals, in SOAs. Optical spectrum of the Manchester signals is also two times that of the NRZ signals. As a result, optical spectral efficiency and dispersion tolerance of the Manchester signals are degraded greatly. It has been known that the use of optical duo-binary modulation format is an effective means in achieving narrow signal optical spectrum. A novel dual-input Mach-Zehnder modulator was used in conjunction with a wavelength modulation scheme to remove the cross-gain modulation in the semiconductor optical amplifier (SOA) [3].

Keywords: Cross-gain modulation (XGM), Wavelength Division Multiplexing (WDM), differential phase-shift keying (DPSK), semiconductor optical amplifier (SOA), Optical Signal to Noise Ratio (OSNR).

1. INTRODUCTION

Cross Gain Modulations is one of the non-linear effects that can arise inside a semiconductor optical cavity (as a SOA). It takes place when a high power signal (called Pump) is injected into the SOA, depleting most of the carriers present in the active region when it is amplified, if we simultaneously inject a low power signal (Probe) in the SOA it will be attenuated due to the absorption of the carriers. In the case where pump and probe are binary signals modulated in amplitude, the XGM causes that the output probe signal can be interpreted as the logic function (*Probe*) AND (*NOT (Pump)*), this response has been extensively referenced in the literature [4]. Compared with erbium-doped fiber amplifiers (EDFAs), semiconductor optical amplifiers (SOAs) offer a number of advantages for high-capacity transmission systems. These advantages include compactness, low cost, low power consumption, and ease of integration with other devices, and ultrawide gain spectrum. However, SOAs suffer from cross-gain modulation (XGM), which results in crosstalk penalty for multiplexed signals [5].

This is due to its relatively low saturation energy and gain recovery time comparable to the bit period. Conventional ON-OFF-keyed (OOK) signals, such as return-to-zero (RZ) and non-return-to-zero (NRZ) formats, have long consecutive “1” or “0” bit, which has high randomness. Hence, they are susceptible to XGM effect in SOAs. Recently, the wavelength modulation technique is proposed to reduce the XGM effect of the NRZ signal in SOAs [3]. However, dispersion in the transmission system limits the effectiveness of this technique. The RZ differential phase-shift keying (RZ-DPSK) signal has periodical amplitude and data is represented in phase only.

2. SYSTEM DESIGN & IMPLEMENTATION

This Paper provides a complete description of the design of the system that how to compare the cross gain modulation (XGM) in SOAs using OOK, Manchester, Duo-binary and DPSK signals using optical communication systems design software (Optisystem 7.0).

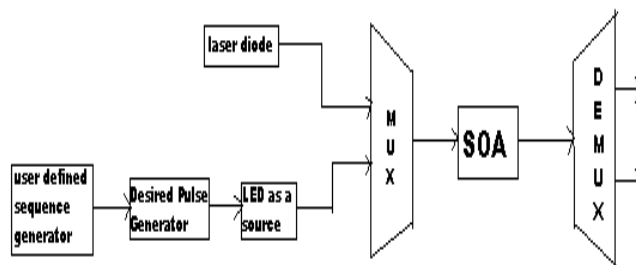


Fig. 1. System Design

Here I designed the above system using optisystem software for OOK, DPSK, Duo-binary & Manchester signal format as a desired pulse generator. Here I am multiplexing high power signal (pump signal) at 1550 nm wavelength from laser diode & low power signal (probe signal) at 1540 nm from LED. Then this multiplexed signal is transmitted through SOA (Semiconductor Optical Amplifier) and then demultiplexes it. Here I am checking the power level of low power signal (probe signal) & high power signal (pump signal) at SOA input & output to check for XGM effect.

3. RESULTS & ANALYSIS

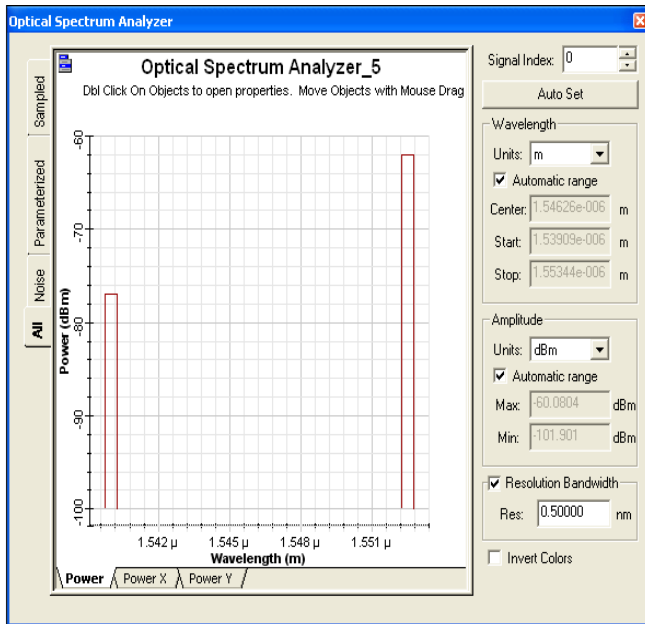


Fig. 2. SOA output shown in spectrum analyzer for RZ

For RZ signal format, I am getting the probe signal power of -74 dBm and pump signal power of -61.844 dBm. Here we are able to see that there is a 30 dBm improvement in power level.

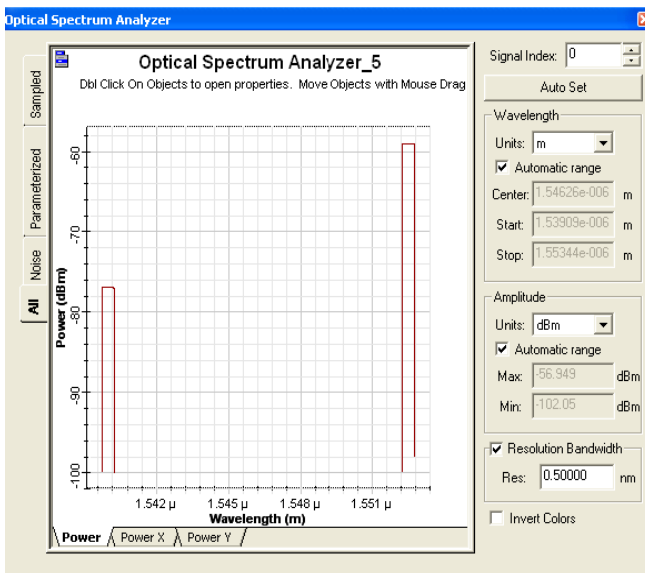


Fig. 3. SOA output shown in spectrum analyzer for NRZ

For NRZ signal format, I am getting the probe signal power of -73 dBm and pump signal power of -59.929 dBm. Here we are able to see that there is a 3 dBm improvement in power level of pump signal and 1dBm improvement of

power level of probe signal compared with RZ signal format.

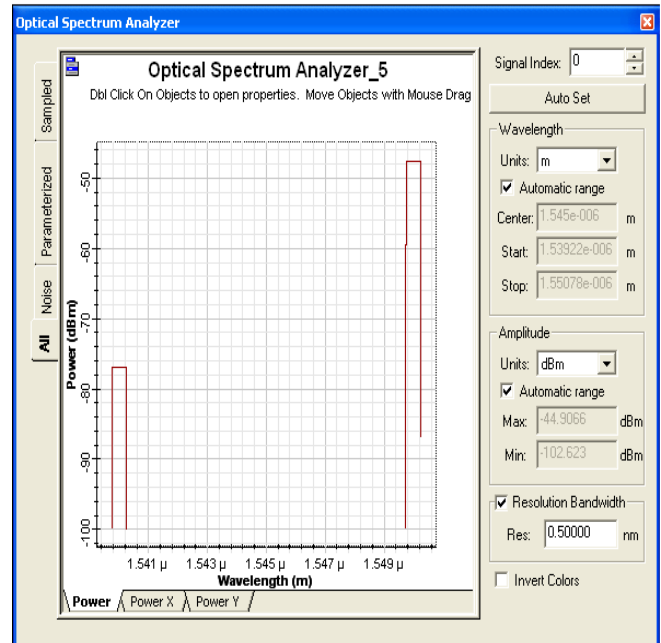


Fig. 4. SOA output shown in spectrum analyzer for DPSK

For DPSK signal format, I am getting the probe signal power of -76 dBm and pump signal power of -47.525 dBm. Here we are able to see that there is a 11 dBm improvement in power level of pump signal and 3 dBm decrement of powerlevel in probe signal compared with NRZ signal format.

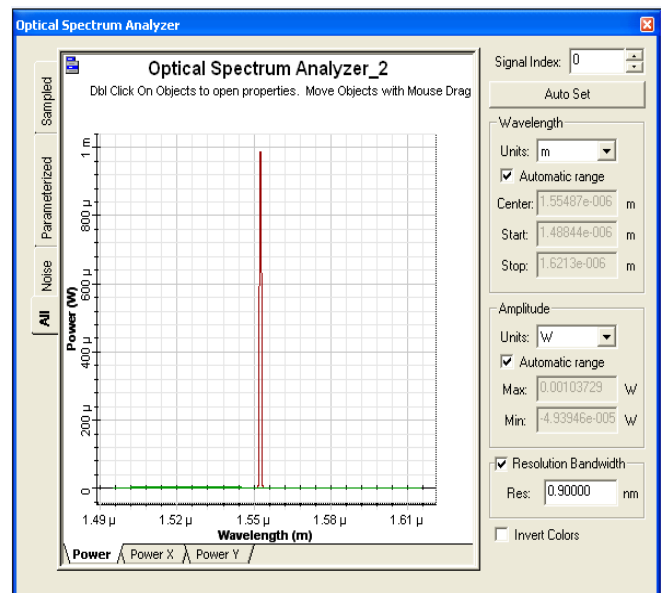


Fig. 5. SOA output shown in spectrum analyzer for Duobinary

For Duo-binary signal format, I am getting the probe signal power of -56 dBm and pump signal power of 0.242 dBm. Here we are able to see that there is a 47 dBm improvement in power level of pump signal and 20 dBm improvement of powerlevel in probe signal compared with DPSK signal format.

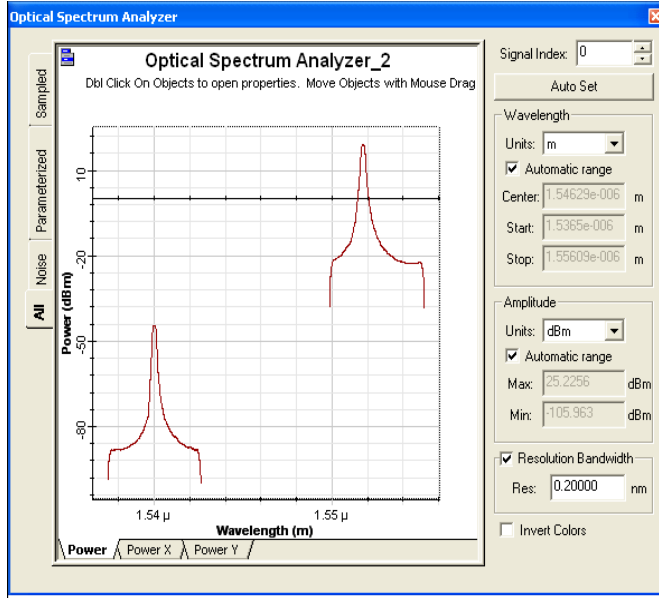


Fig. 6. SOA output shown in spectrum analyzer for Manchester

For Manchester signal format, I am getting the probe signal power of -44 dBm and pump signal power of 19.777 dBm. Here we are able to see that there is a 19.5 dBm improvement in power level of pump signal and 12 dBm improvement of powerlevel in probe signal compared with Duo-Binary signal format.

Following table shows that Manchester signal has highest power for both the signals, Pump as well Probe signal, compared to all. So, it is used to remove cross gain modulation (XGM) effect.

Table I: Cumulative Results

Sr. No.	Modulation Format	SOA Output Power (Pump signal)		SOA Output Power (Probe signal)
		dBm	Watts	dBm
1	RZ	-61.844	654.03 pw	-74
2	NRZ	-58.929	1.28 nw	-73
3	DPSK	-47.525	17.68 nw	-76
4	DUO-BINARY	0.242	1.057 mw	-56
5	MANCHESTER	19.777	95.022 mw	-44

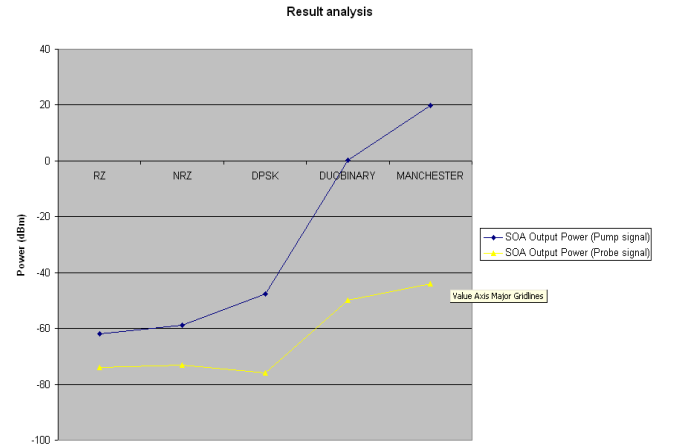


Fig. 7. Result analysis using line chart

4. CONCLUSION

From the design and the implementation of the system presented in this document it can be concluded that this system could solve many of the challenges that are faced by cross gain modulation effect (XGM) in semiconductor optical amplifiers (SOAs) in optical communication system. Because of the flexibility and ease of designing by using the optical communication system design software, presented system could be adapted for any application for long haul optical communication in which semiconductor optical amplifiers are used. Here the cross gain modulation (XGM) effect in SOAs using OOK, Manchester, duo-binary & DPSK coding is investigated and compared. For future this system can solve many technological challenges.

5. ACKNOWLEDGEMENT

Among the countless people to whom I am grateful for accompanying me during these study, a first heartfelt thank goes to my guide Prof. Hemant Purohit, who followed and directed me for this paper writing. I would like to express my deep sense of gratitude to Prof. O. P. Vyas, (Dean Engineering) for guiding me with attention and care. I would like to express my sincere gratefulness to dear GOD, my parents and my dear friends and all those people who have helped me directly and indirectly.

REFERENCES

- [1] H. K. Kim and S. Chandrasekhar, "Reduction of cross-gain modulation in the semiconductor optical amplifier by using wavelength modulated signal," *IEEE Photon. Technol. Lett.*, vol. 12, no. 10, pp. 1412–1414, Oct. 2000.
- [2] D. T. Schaafsma and E. M. Bradley, "Cross gain modulation and frequency conversion crosstalk effects in 1550 nm gain clamped semiconductor optical amplifiers," *IEEE Photon. Technol. Lett.*, vol. 11, no. 6, pp. 727–729, Jun. 1999.

-
- [3] Y. Dong, Z. Li, C. Lu, Y. Wang, Y. J. Wen, T. H. Cheng, and W. Hu, "Improving dispersion tolerance of Manchester coding by incorporating duobinary coding," *IEEE Photon. Technol. Lett.*, vol. 18, no. 16, pp. 1723–1725, Aug. 1, 2006.
- [4] Asier Villafranca, Ignacio Garcés, Miguel Cabezón, Juan José Martínez, David Izquierdo, José Pozo, "Multiple-Bit All-Optical Logic Based on Cross-Gain Modulation in semiconductor amplifier", *ICTON 2010*.
- [5] Zhaohui Li, Yi Dong, Chao Lu, Yang Jing Wen, Yixin Wang, Weisheng Hu, and Tee Hiang Cheng, "Comparison of Cross-Gain Modulation Effect of Manchester-Duobinary, RZ-DPSK, NRZ-DPSK, RZ and NRZ Modulation Formats in SOAs", *IEEE Photon. Technology Lett.*, vol. 18, no. 24, dec 15, 2006.
- [6] Francesco Marino, Luca Furfaro, and Salvador Balle, "Cross-gain modulation in broad-area vertical-cavity semiconductor optical amplifier", *APPLIED PHYSICS LETTERS* 86, 151116 (2005).
- [7] A. Bilenca R. Alizon, V. Mikhelashvili, D. Dahan, G. Eisenstein, R. Schwertberger, D. Gold, J. P. Reithmaier, and A. Forchel, "Broad-Band Wavelength Conversion Based on Cross-Gain Modulation and Four-Wave Mixing in InAs-InP Quantum-Dash Semiconductor Optical Amplifiers Operating at 1550 nm", *IEEE PHOTONICS TECHNOLOGY LETTERS*, VOL. 15, NO. 4, APRIL 2003.
- [8] P. S. Cho and J. B. Khurgin, "Suppression of cross-gain modulation in SOA using RZ-DPSK modulation format," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 162–164, Jan. 2003.
- [9] Jan Lamperski, "Cross –Gain Modulation Techniques for All Optical Wavelength Conversion", *XIV Poznań Telecommunications Workshop - PWT 2010*.

Retrieval of Target Velocity using Doppler's Effect Phenomena by Comparative Study of L, C and X Band Radar

Garima Sharma¹, Nishank Agarwal², Aditya Sharma³, Sudhir Kumar Chaturvedi⁴,
Pavan Kumar Nanduri⁵, Ugur Guven⁶

^{1,2,3}*B. Tech Avionics Engineering Students,
Department of Aerospace Engineering
University of Petroleum & Energy Studies, Dehradun, India
garimasharma1209@yahoo.com*

^{4,5,6}*Faculties, Department of Aerospace Engineering
University of Petroleum & Energy Studies, Dehradun, India
sudhir.chaturvedi@ddn.upes.ac.in*

Abstract: This paper address the estimation of target velocity by RADAR system datasets. Radar is having the significant potential for estimating the various parameters such as size, shape, direction, and speed for the various types of targets. If the targets moves with certain velocity either away from radar or it approach the radar, the change in signal phases occurs. The phase difference between transmitted signal and echoes can be useful for studying the concept of the velocity vector determination under the given power ratio for the transmission and received pulses. Target under stationary conditions there won't be any change in the signal parameters.

The paper focuses on retrieval of the velocity of moving targets using the standard Doppler's effect phenomena and the results will be compared with the though transform algorithms. Doppler's effect provides the estimation of radial velocity of the target in horizontal and vertical direction from the body frame or axis. The Hough transform is very useful in this case; its application contributes to retrieve the target parameters under the low signal-to-noise ratio. The relation between and radial velocity and phase will be developed to obtain the expected result of the problem. The result will be implemented over the three band of frequencies such as L, C and X band to obtain the highly effective comparison result. The combined study of both the algorithm provides a good estimate of result and it may also provide to design an integrated system for the retrieval of velocity for the moving targets using some other navigation systems such as Long Range and Navigation (LORAN), Distance Measuring Equipments (DME) to get the better security and safety system used for defence technologies.

Index Terms: Radar, Doppler's Effect, Hough Transform, Frequency Band, Target Velocity

1. INTRODUCTION

Recently, mathematical methods for extraction of useful data about the behavior of observed targets by mathematical transformation of received signals are being widely used for

design of new highly effective algorithms for processing radar data. Modern methods for target detection and trajectory parameters estimation, which use mathematical transformation of received signals, allow new highly effective algorithms for radar signal processing to be designed. As a result, extremely precise estimates of moving target parameters can be obtained in conditions of very dynamic radar environment. In conventional radar systems, the radial velocity of targets in the radar directional beam is measured using the Doppler Effect. In each range resolution cell, velocity measurements are made by transmitting a pulse train towards a target over a very short period of time, and measuring relative target movement between each pulse. The number of pulses used is usually known as packet size and the frequency in which they are transmitted as pulse repetition frequency. The Doppler radar processes a train of received pulses determining the average phase-shift between successive pulses within a pulse packet. This is typically done by means of a 1D Fast Fourier Transform (FFT), which is performed independently for each range resolution cell, using all pulses within a packet. The radial velocity of the target is evaluated based on knowledge of the radar frequency, speed of light, pulse repetition frequency and average Doppler phase-shift.

An alternative approach for velocity estimation of targets moving towards or down radar can be realized using the Hough detector proposed by Carlson (1). An improved system concept involves a processing method, which allows preceding data to help in target detection. For e.g. The Hough Transform techniques detect the existence of tracks using batch of frames (3). In paper (2) two techniques for radial velocity estimation that uses Hough and Doppler transform are considered and compared based on some parameters of surveillance radar.

2. MATERIAL AND METHOD

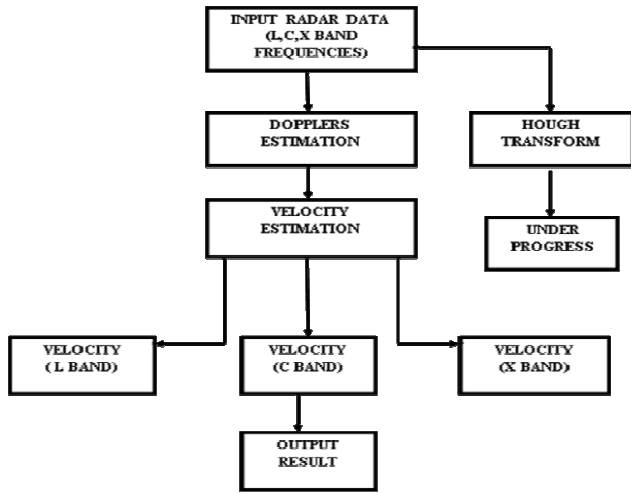


Fig. 1. Flowchart for comparative study for the target velocity estimation using RADAR

Figure 1 shows the basic flow chart for the retrieval and comparison of the velocity of target by means of Doppler's effect and Hough transform. Hough transform work is under progress (not presented in this paper). Different band such as L (2.56GHz), C (4.5GHz) and X (10GHz) bands were selected for this study in order to study the comparison result. A feasible technique for separating the received signal from the transmitted signal when there is relative motion between radar and target is based on recognizing the change in the echo-signal frequency caused by the Doppler effect. It is well known in the fields of optics and acoustics that if either the source of oscillation or the observer of the oscillation is in motion, an apparent shift in frequency will result. This is called Doppler effect (4, 5).

In conventional radar systems the radial target velocity is measured using the Doppler Effect. The principal structure of a Doppler estimator is shown in figure. The incoming radio frequency signal is demodulated down to a center frequency of zero prior to pulse compression and Doppler filtering. This is down to reduce computational burden, since the demodulated signal can be down sampled to reduce the amount of data needed for storage. The demodulated signal is usually referred as complex envelope or IQ-data, where I-data is a real part and Q-data is an imaginary part of a complex envelope.

After pulse compression, the complex amplitude of all pulses received from a target is processed into Doppler velocity filters which are used to determine velocity. The envelope at the output of a filter with maximal envelope is compared with an adaptive Constant False Alarm Rate (CFAR) detection threshold. If this threshold is exceeded, the radial velocity of the target is evaluated by equation 1.

$$V_{Target} = \frac{\lambda F_{PRF} n}{2 N_{FFT}} \text{ where } F_{PRF} < \frac{c}{2 R_{max}} \quad (1)$$

Where, F_{PRF} = Pulse Repetition frequency, N_{FFT} = No. of pts in FFT transform, R_{max} = Maximal Unambiguous Range, n = channel no.

3. RESULTS AND DISCUSSION

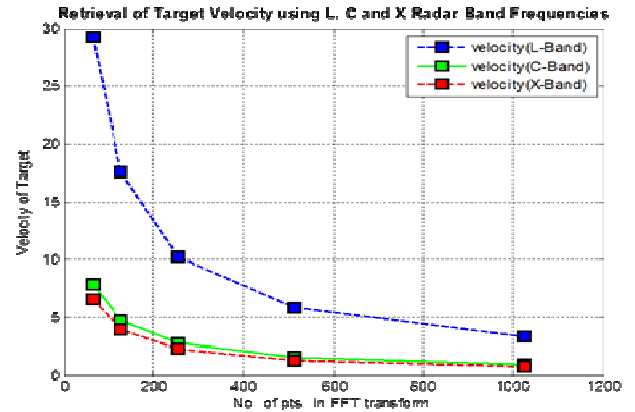


Fig. 2. Comparative result of target velocities in L, C, and X band radar frequencies.

Figure 2 shows the comparative result for the L,C and X band frequencies. The simulation has been carried out in MATLAB programming with selecting the different number of FFT points. As per the equation 1, it can observe that velocity of target decreases if the number of FFT points is increases. The final result is presented in Table 1.

Number of FFT depends on the number of channels used for the transmission of each signals from radar transmitter, the higher deviation in velocity can be seen for the L band in comparison with C and X band frequencies because of its shorter frequencies and higher wavelengths. C and X band velocity does not show the much deviation, hence these bands are useful for the estimation of target speeds in defence technologies.

4. TARGET VELOCITIES FOR L,C AND X BAND RADAR DATASETS

NFFT	Velocity (L-Band)	Velocity (C-Band)	Velocity(X-Band)
64	29.2969	7.8125	6.5104
128	17.5781	4.6875	3.9063
256	10.2539	2.7344	2.2786
512	5.85943	1.5625	1.3021
1024	3.2959	0.8789	0.7324

5. CONCLUSION

A technique for retrieving the velocity using radar has been carried out in this paper. The Doppler's effect phenomena has been considered to estimate the velocity for L,C and X band frequencies with the different channel transmission from the radar transmitter. The various velocity has been obtained using the 3 band frequencies of microwave spectrum and result shows the less percentage of error in case of C and X band frequencies due to less attenuation losses in to atmosphere and target interaction. The work is under in progress and in future we will implement the Hough transform to estimate the velocity for the same band frequencies. The result will be used for validation purpose.

6. ACKNOWLEDGMENT

The author (s) would like to thank University of Petroleum & Energy Studies, Dehradun for providing the opportunity to write this research article.

The Corresponding author also would like to thank Dr. C.S.

Yang, Korea Ocean Research & Development Institute, Korea (Republic of) and Dr. P. Shanmugam, Indian Institute of Technology Madras, India for providing his support towards this work.

REFERENCES

- [1] B. Carlson, E. Evans, and S. Wilson, "Search radar detection and track with the hough transform" IEEE transaction of remote sensing, vol. 30, pp 102-124, 1994.
- [2] V. Behar, L. Doukovska, C. Kabakchiev, and H. Rohling, "Comparison of Doppler and Hough Target Velocity Estimation Techniques", Proc. of the International Radar Symposium, Cologne, Germany, pp.157-162, 2007.
- [3] L. Doukovska, and D. Angelova. "Comparative Analysis of Two Techniques for Moving Target Velocity Estimation", Proc. of the 7th European Radar Conference, 2009
- [4] M. Skolnik, Introduction to Radar System..McGraw Hills International Education: 3rd editions.
- [5] L. Doukovska, "Combined Doppler-Hough method for velocity estimation". Proc. of International Radar Symposium, Germany, 677-688, 2009.

Enhanced QoS Through Traffic Pacified Handoff Algorithm in Wimax 16m Networks

D. Karunkuzhali¹, D.C.Tomar²

¹Research Scholar, Department of CSE, Sathyabama University, Chennai, India
karunkuzhali@gmail.com

²Professor, Department of CSE,
Shree Motilal Kanhaiyalal Fomra Institute of Technology, Chennai, India
dctomar@gmail.com

Abstract: The IEEE 802.16m standard for Advanced mobile broadband wireless access provides a seamless application connectivity to other mobile and IP networks like UMTS, LTE and WLAN which are having great difference in terms of data transmission rate, Coverage, cost and supporting of service types. Users' mobility is a major factor which directly affects the performance of the communication network. In this paper, we propose an improved Traffic pacified handoff algorithm for the buffer management and to enhance the quality of service features in the air interface standard of IEEE 802.16 m. The standard aims to reduce overhead, improved coverage through optimized parameterization and excessive security measures. These parameters define the set of services that are provided for protection and they are agreed upon at the time of security association establishment. Optimization in the threshold value is achieved through the calculation both SNR and delay in a combined manner to evaluate the heuristic handoff for a given scanning interval. We evident that our proposed system framework performs efficient handoff management with effort from base station scheduler and subscriber station scheduler. These processes incur least bandwidth under utilization thereby reducing the transmission delay under heavy traffic condition with QoS features conciliation.

Keywords: Transmission delay, handoff, optimized parameterization, buffer range, Quality of Service.

1. INTRODUCTION

The IEEE 802.16 standard [1,2,3,4] is a promising standard for next-generation broadband wireless access networks. It provides last mile solution and supports high data-rate transmissions, extensive-area coverage, and high-speed mobility for users with multimedia applications. It plays an important role in shaping 4G mobile networks by supporting IMT-A requirement by updating its IEEE 802.16 standards to meet the requirements of next generation mobile networks targeted by the cellular layer of IMT-Advanced.

The IEEE 802.16m standard for Advanced mobile broadband wireless access provides a seamless application connectivity to other mobile and IP networks like GSM, WCDMA, UMTS, LTE and WLAN which are having great difference in terms of data transmission rate, Coverage, cost

and supporting of service types. Emerging multimedia services like immersive environments, IPTV applications, video conferencing, and 3D virtual world requires reliable communication even in high mobility in heterogeneous network environment, denser area and in cell edges.

The key features of Wimax 16m network systems include (1) increased spectral efficiency and bandwidth (2) improved cell edge (3) performance and mobility support (4) reduced control and user plane latency and (5) reduced handoff interruption time. Orthogonal frequency division multiple access (OFDMA), described as a key technology for Wimax physical layers, is used to adjust channel bandwidth and to allocate subscriber station (SS) subcarriers according to channel state. OFDMA also allows multiple SSs to use various subcarriers to simultaneously transmit OFDM symbols, so called SOFDMA (Scalable-OFDMA). In a base station (BS), all OFDMA subcarriers are divided into groups (known as subchannels) that are allocated to different SSs with matching bandwidth and quality of service (QoS) requirements [4].

SOFDMA (Scalable-OFDMA) has several advantages to be considered for the Wimax networks, such as flexibility of allocating subcarriers to users; high spectral efficiency, low receiver complexity, and simple implementation by fast Fourier transform (FFT) and inverse FFT are also considered. In addition, adaptive modulation and coding (AMC) technology allows SOFDMA PHY to facilitate data transmission in a high mobility environment, and makes wireless resources fully utilized. This AMC technology uses the CQI (Channel Quality Information) channel to determine the appropriate Modulation and coding scheme. The channel quality is determined by the instantaneous received Signal-to-Noise Ratio (SNR). To determine the mode of transmission (i.e., modulation level and coding rate), an estimated value of SNR at the receiver is used. To avoid possible transmission error, no packet is transmitted when the SNR is less than the predefined threshold Value [9,10,11,12].

A. Mobility Management in Wimax 16m

Mobility management is a main challenge in the Wimax network. Mobility management Techniques that support user movement within and between different networks are defined. Exploring the handoff performance should not confine to mere air interface investigation since many operations are performed inside the access and core network.

Usually a handoff is understood as a change of physical connection point through which the terminal communicates with network services. From wide perspective handoffs may be split into two groups: horizontal HO and vertical HO. In the horizontal HO network technology remains the same, whereas the latter is an inter-technology HO type. Media Independent Handoff (MIH) is a vertical HO method defined in 802.21 specifications.

In general, handoff is performed in three phases: (i) network discovery, (ii) hand decision and (iii) handoff execution. The first phase is network discovery which is the process by which a Mobile Node finds/detects the target network. In the second phase the decision for handoff is made. This phase represents the criteria function, based on which decision is made, taking into account different parameters. Several suggestions for trigger and optimization criteria are presented in [5] and may be summarized as follows:

- Received Signal Strength (RSS) or SNR (e.g. user uses the network with the best available signal);
- QoS parameters in the network (e.g. some applications require a high level of QoS support);
- bandwidth of the target network (e.g. user uses the network with the broadest bandwidth);
- power consumption (e.g. some network interfaces require higher power, which can lead to greater battery consumption);

2. RELATED WORK

In Wimax 16m, Handoff is characterized as hard handoff (also known as break-before-make) and soft handoff. Soft handoff consists of Macro Diversity Handoff (MDHO) and Fast Base Station Switching (FBSS). Soft handoff improves the Quality of Service (QoS) performance while adding complexity and overhead to the system. Handoff process is initiated based on the measurement of the signal strengths received by Mobile Station (MS) from multiple Base Stations (BSs). To describe the signal quality without taking the receiver into account, the optical SNR (OSNR) is used [11]. The OSNR is the ratio between the signal power and the noise power in a given bandwidth [9]. Most commonly a reference bandwidth of 0.1 nm is used. This bandwidth is independent of the modulation format, the frequency and the receiver. For instance an OSNR of 20dB/0.1 nm could be

given, even the signal of 40 GBit DPSK would not fit in this bandwidth [10]. OSNR is measured with an optical spectrum analyzer that minimizes the noise overheads.

Handoff occurs frequently because of the channel traffic load and the wireless environment that causes channel fading and shadowing. Most reported algorithms depend on various handoff criteria such as signal to noise ratio (SNR) or the received signal strength indicator (RSSI). These algorithms may be divided into three categories. In the first category, handoff decision is initiated when the received signal strength of the serving BS is lower than the received signal strength of target BS. Repeated and unnecessary handoffs may occur even if the MS receives a signal with acceptable SNR, which affects the performance of the system and degrades QoS of the connection.

In the second category, the decision is based on relative signal strength and the threshold. This method may prevent the repeated handoffs between two BSs, which occur in the algorithms in the first category when the MS reaches the cell boundary. However, an optimization of the threshold value is required because choosing a large threshold reduces the handoff attempts and, consequently, delays the handoff initialization and degrades the connection quality. The third category is based on the relative signal strength with a threshold and a margin. The handoff is initiated only when the current received signal strength from the serving BS is lower than a certain threshold and the SNR of the target BS is higher than the SNR of the serving BS. In this case, the repeated handoffs are prevented and the coverage area of the BSs is maximized. The drawback of this method is the optimization overhead of both the handoff threshold and the margin: low threshold causes degraded connections due to late handoff while high threshold causes premature handoff. Both affect the coverage and the system throughput.

The Dual Triggered Handoff algorithm[13] is a hybrid of MS initiated and the BS initiated capacity handoffs in the 802.16e, a version to deliver service across many more sub-channels. Here each subscriber is linked to a number of sub-channels that prevent multi-path interference. It means that it permits an arriving MS to perform the handoff to a target BS that already operates at zero free capacity. The handoff threshold value specifying the minimum difference between the SNR of a neighboring BS and SNR of the serving BS before triggering a handoff to replace the serving BS with a neighboring BS is to be reduced or analyzed further to produce an optimum solution. And in order to provide network security level number of global comprehensive algorithms is proposed to align the traffic levels in a network with respect to the capacity of the regions of handoff. And for the purpose of detecting traffic pattern the Packet Analyzer were used and imported to other modules in a network other than discriminating it.

3. PROPOSED TRAFFIC PACIFIED HANDOFF ALGORITHM

In our proposed system the process of handoff is enhanced with effort from optimized parameterization concentrates over the handoff favoring the QoS features in the air interface standard of IEEE 802.16m. The standard aims at Reduced overhead, improved coverage through optimized parameterization and excessive security Protection. These parameters define the set of services that are provided for protection and are they are agreed upon at the time of security association establishment. It includes the request and response of both the MS and the BS in order to measure the parametric values. And the respective threshold values used is also found to be optimum in handoff operations. Optimization in the threshold value is achieved through the calculation both SNR and delay in a combined manner to evaluate the heuristic handoff for a given scanning interval.

These methodologies were followed through a sequence of process that initiates with module configuration. In this process modularity of the system is checked for its identity and preference in the whole handoff process. Then the point beyond which there is a change in the manner a program executes in particular, an error rate above which the system ends on the assumption that a failure has occurred is calculated through assigning the values after scanning the system. Then the signal to noise ratio where the ratio of the amplitude of a desired analogy or digital data signal to the amplitude of noise in a transmission channel at a specific point in time is calculated with respect to service station which enables the connection with that of delay factor which is a function of the packet's length and has nothing to do with the distance between the two nodes. This delay is proportional to the packet's length in bits. A switch using store-and-forward transmission will receive the entire packet to the buffer and check it for CRC errors or other problems before sending the first bit of the packet into the outbound link. Thus store-and-forward packet switches introduce a store-and-forward delay at the input to each link along the packet's route.

A. Handoff Probability

Let the number of channels be denoted by N. Handoff Probability can be defined as the probability that a call needs atleast one more handoff during its remaining life time. It characterizes whether an ongoing call completes its session in the current cell or not.

The steady state probability of hand off is given by,

$$P(j) = \frac{(\lambda_C + \lambda_H)^j}{j! \mu^j} \cdot P(0), [j = \text{sub-cells}] \quad (1)$$

Where the initial probability of $P(0)$ is given by,

$$P(0) = \left[1 + \sum_{j=1}^N \left(\frac{\lambda_C + \lambda_H}{\mu} \right)^j \cdot \frac{1}{j!} \right]^{-1} \quad (2)$$

The channels with identical cells have same statistical behaviour and the traffic can be of Poisson distribution, in case of that if N is large then an approximate independent model of traffic level can be figured out. Mean (or) Cell holding time of a station of the transmission is given by μ .

$$\frac{\lambda_C}{\lambda_H} = \frac{P_H(1-P_B)}{1-P_H(1-P_H)} \quad (3)$$

Where k = better buffer capacity. λ_H and λ_C denotes the out of cell and within cell traffic handoff rates and P_H and P_B denotes the blocking probabilities of handoff block and new cell block.

On substituting (3) in (2) we get

$$P(0) = 1 + \sum_{j=1}^N \left[\frac{\mu(1-P_H) \cdot j!}{(1+P_H[PH-PB])} \right] \quad (4)$$

Now the probability value of $P(j)$ can be given as,

$$P(j) = \frac{(1+P_H[PH-PB]) + \sum_{j=1}^N \left(\frac{\mu(1-P_H) \cdot 1-j}{((1+P_H[PH-PB]))^{1-j}} \right)}{\mu(1-P_H) \cdot j!} \quad (5)$$

And the handoff probability is given by,

$$P_h(k) = \frac{\sum_{i=0}^{m-1} \binom{k+i-1}{i} \left(\frac{\eta}{\alpha+\eta} \right)^i \left(\frac{\eta}{\alpha+\eta} \right)^k}{\sum_{i=0}^{m-1} \binom{k+i-2}{i} \left(\frac{\eta}{\alpha+\eta} \right)^i \left(\frac{\eta}{\alpha+\eta} \right)^{k-1}} \quad (6)$$

And similarly,

$$P_{ho} = Prob \left\{ \bigcup_{i=2}^j [R_i > R_l + h] \right\}$$

$$P_{ho} = 1 - \prod_{i=2}^j Prob[R_i \leq R_l + h] \quad (7)$$

In equ(4) P_h value can be applied from (7), specifying the heuristic value and satisfying received signal strength condition.

The above equation implies that, Optimal handoff probability can also be obtained with respect to the heuristic value of optimal h . Here R_i denotes the RSS (Received Signal Strength) of the target BS and R_l the respective serving BS. RSS values can also be explained with RSS Index (RSSI) index value. ' η ' being the cell residence time can be used in calculating the call to mobility factor of, $\alpha = (\mu/\eta)$. (8)

From the above the delivered buffer load can be obtained as

$$BL = \frac{\text{Maximum Traffic Rate}}{1-P(0)} \quad (9)$$

$P_{(0)}$ = initial Probability;

In case of initial probability the traffic level rates are considered to be a constant.

Time step values(T_s) can be user specified matching the application in TDD. The slots per frame of the TDD can thus be calculated as $BL * T_s$.

B. Handoff Delay

When the buffer is empty, the maximum number of channels which can be used to transmit new cells in n ($n \leq N$). Consequently all remaining channels,

$h = N - n$, can be occupied by handoff calls. However when the buffer is not empty, 'b' more channels are allocated to handoff calls in order to reduce the delay of handoff calls.

The evaluated time period (T) can be given as,

$T = b/v$. $b \Rightarrow$ average interval period. $v \Rightarrow$ velocity of the Mobile Node.

Hence the average handoff delay can be given as $\delta_{avg} = T/2$. While considering the moment of transmission of REQ message in "Handoff Process", the delay(τ) can be calculated as

$$\tau = (Ct/R).T_{UL} + mod(Dc,R).T_d(10)$$

where

T_d : Transmission Duration.

T_{UL} : Uplink Transmission duration.

Ct : Contention transmission value for server.

R : Initial Ranging interval area.

C. Indexing

Based on the buffer capacity value encompassing the load and traffic capacity, indexing can be done in 'j' ($j \in R$) ranges of the given predefined area. The progression of localization is initiated here as a set of base stations that are grouped together to optimise signalling. Typically, tens or even hundreds of base stations share a single Base Station Controller (BSC), the intelligence behind the base stations. The BSC handles allocation of radio channels, receives measurements from the mobile phones and controls handoffs from base station to base station. The transformation of packets is made with the dynamic buffering thereby favouring noiseless buffering capacity and also reduces the ranging of segmented regions while increasing the gain of signal. The latency delay produce during handoff operation is also considerably reduced.

Traffics in network are best formulated through the security policies of the networking layer. Traffic rate (Packets/Sec)

information should be shared between both the sharing and the shared stations. In Wimax configuration IPSec parameter concerns the network security and aligns the traffic level. Of traffic levelling and controlling during handoff, the categorized (range) area of the network is evaluated in terms of capacity of information handled by the assessed/indexed region. Hence by providing pacified traffic level in the declared assessment region applicable QoS values of delays and buffer capacity is met with our proposed concept in IEEE 802.16m. This handoff control in 802.16m allows the application pacified traffic level in defined range and this whole system is shown in figure 2. Then the vital process of network configuration that works on setting up network components for respective protocol is not trivial for use outside your LAN (Local Area Network). Since so many firewalls and routers exist, it is impractical to give detailed step-by-step instructions suitable for every user. It is important to understand the basics of the respective protocol in order to configure. Next is mainly used for mapping textures to surfaces. Recently, it has become a powerful tool for many applications in mesh processing of parameterisation of network modules, once these parameters satisfies, an optimum value is set to the scanning threshold (T). Measurement of the value of SNR and delay factor of the serving base stations is done it is mentioned as D_{snr} . A circumstance is check of maxmin criteria where threshold value exceeds signal to noise ratio value the process will be redirected to scanning of T again.

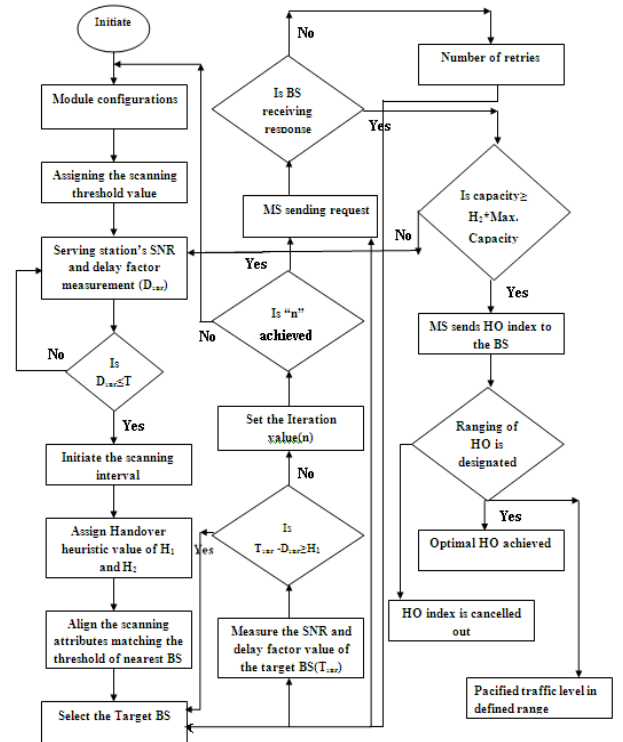


Fig. 1. Algorithm of Traffic based duo triggered Handoff

In other case selection of values from the scanning parameters is done with respect to the T value and an approximate scanning interval (N) is assigned for the breaking up of operation at regular interval of time if needed. Thus selection of target base station is made according the serving base station and as an intermediate procedure redetermination the vital signal to noise ratio and delay factor D_{snr} and T_{snr} is made respectively which ensures the accuracy in handoff hysteresis value. The hysteresis margin is commonly used parameter for elimination of redundant handoffs. This is the dependence of a system not only on its current environment but also on its past environment. This dependence arises because the system can be in more than one internal state H1 for first handoff and H2 for second handoff. The process flows correspondingly only when the difference of T and D_{snr} value exceeds with H1 and provides handoff through selection of target base station. In the other case the process will be sent to calculation of interval again (N). Once the target BS is selected a handoff request is sent from the mobile station MS through the process. At first Station issues association request to new access point then New AP sends the association response to the station and New AP sends multicast handoff request to old AP.

If Old APs association with the station is younger than a handoff hold-down period (e.g. 5 seconds), the Old AP sends disassociation to station. If old association is older, disassociation has no benefit. As a precautionary measure Using an alert to recover from a lost handoff is made by making Old AP discard a frame for a station, which should have responded. Since it queued the frame for the station, it saw no frame from the station (no control, data, or management) then the Old AP includes station in list of potentially lost handoffs in a multicast announce frame (an alert). This is rate limited to prevent announce storms. Finally any AP that has an association with the station regenerates a handoff that needs no response. Once if the target BS receives the response its starts to buffer the incoming response with network capacity calculation to withstand the buffering. If not the system goes of maximum number of retries to receive the response.

*Current capacity of buffer $\geq H_2 * \text{Maximum capacity}$*

In case of satisfying the above mentioned criteria the hand off request from index from MS is sent to the serving BS to confirm the handoff is connected with the valid target BS. Otherwise reconfiguration of buffering capacity is done with can hold the incoming response. Then initial range of Hand off is calculated with respect to the target BS. If the Network entry is successful, then the handoff is made successful with dynamic buffering if not the confirmation process is repeated, following the index cancellation by the serving BS. Thus efficient routing along with improved handoff is achieved by the way of reducing time delay and packet loss

during sending and receiving process at both the base stations thereby we can increase throughput of packet transmission.

4. SYSTEM IMPLEMENTATION

The proposed system is with enhanced QoS is implemented in terms of above mentioned 802.16m with handoff standards. The following flow will show the entire process in terms of codes with vital handoff control mechanism.

Pseudo Code:

- [1] Initiate.
- [2] Process Initiation.
- [3] Locating the base stations, routers and subscriber stations.
- [4] Network configuration.
- [5] Parameterization of the network modules.
- [6] Set an optimum value of scanning threshold, let it be T.
- [7] Measure the value of SNR and delay factor of the serving base stations, D_{snr} .
- [8] for 1:n iterations
- [9] if ($D_{snr} \leq T$)
- [10] Goto step 12
- [11] else
- [12] Goto step 6.
- [13] Scanning Process
- [14] Based on the value of T, select the values of scanning parameters.
- [15] Provide an approximate value of scanning interval (N).
- [16] In accordance to the serving base station select the target base station.
- [17] Now calculate both the values of SNR and delay factor of serving base station and target base station i.e., D_{snr} and T_{snr} .
- [18] Set the values of handoff hysteresis.
- [19] H_1 - First handoff threshold hysteresis, with respect to the scanning process.
- [20] H_2 - Second handoff threshold hysteresis, with respect to the scanning process.
- [21] for 1:n iterations
- [22] if ($T_{snr} - D_{snr} \geq H_1$)
- [23] $SNR_{maxDT} - SNR_{DS} \geq H_1$
- [24] Goto step 27
- [25] else
- [26] Goto step 16.
- [27] Handoff Process
- [28] The available base stations (BS) select the target BS.
- [29] On identifying the target BS send the Hand -Off request from the MS(Mobile Station).
- [30] Is the target BS received the Hand Off response
- [31] If(yes)
- [32] Proceed the buffering process
- [33] else

- [34] Opt for the maximum number of retries to receive the response.
- [35] Calculate the buffering capacity of the network
- [36] If(current capacity of buffer $\geq H_2$ * Maximum capacity)
- [37] Goto step 39
- [38] Else goto step 32.
- [39] Now the hand off request from index MS is sent to the serving BS to confirm the handoff is connected with the valid target BS.
- [40] The Initial range of Hand off is calculated with respect to the target BS.
- [41] If the Network entry is successful, then the handoff is made successful with dynamic buffering.
- [42] Or else the confirmation process is repeated, following the index cancellation by the serving BS.
- [43] End.

5. EXPERIMENTAL STUDY

We have presented more accuracy in terms of efficiency statistically. The consideration is made in order to show the advancement of every parameter like time, throughput, bandwidth, SNR, packet delay, dynamic buffering and packet loss rate.

In our proposed system during the data transmission the exactness is calculated with that of number of packets receiver in the mean interval of time the following graphs indicates the analysis resultant of 802.16m with handoff showing the measurement of previous 802.16e standard. The initial process of network simulation is done through Network Simulator 2 to validate the proposed algorithm. To simplify the simulation scenarios, each BS is assigned a Media Access Control (MAC) address (BS ID) corresponding to its name: MAC i for BS $_i$, $i = 0, 1, 2$, and 3. The OFDMA frame has 512 subcarriers and 5 ms duration. Each MS nodes has a constant downlink traffic flow of 64 kbps to a server throughout the uplink of the target BS. The handoff messages are negotiated through the backbone links between the serving BS and the neighbouring BSs. The network topology has identical object's attributes configurations for all simulation scenarios. Each BS initially has 0.7 Msps free upload link capacity. In our system we have considered a network with nine individual nodes, two routers and a gateway which connects both the routers with that of the server. the BS attribute specifications of scanning and handoff operations where the interleaving attribute is set to 240 to assure maximum throughput with a single row count Row0. This row consists of scanning threshold of 50dB and the handoff parameter where the retransmission time is set to 30 seconds with 0.4 as the handoff threshold hysteresis. QoS attribute configuration node defines the attribute configuration details for protocols supported at the IP layer. The PHY and MAC layers have been configured for the same scenario and their

important parameters where BS station MAC address is assigned on the basis of distance is processes and the buffering ranging is specified as 64 KB with retransmission time interval of .1 seconds. A 20 MHz bandwidth with 5 GHz base frequency is configured to study the effect of heavy traffic on each QoS class with different scheduling mechanism.

When the time measurement is at equal interval of seconds in both the cases and the throughput in our communication networks, like 802.16 is the average rate of successful data delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node. The throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second or data packets per time slot.

Table 1. Scanning and Handoff Parameters.

Scanning & Handoff parameters for BS & SS	
Scanning Threshold (dB)	50
Scan Duration (N) (Frames)	5
Interleaving Interval (P) (Frames)	240
Scan Iteration (T)	10
Scan Request Retransmission (ms)	50
Maximum Scan Request Retransmissions	8
Handoff Threshold Hysteresis (dB)	6.0
MS Handoff Retransmission Timer (ms)	30
Maximum Handoff Request Retransmissions	6
Multitarget Handoff Threshold Hysteresis (dB)	0.0

A. Performance evaluation of Proposed IEEE 802.16m

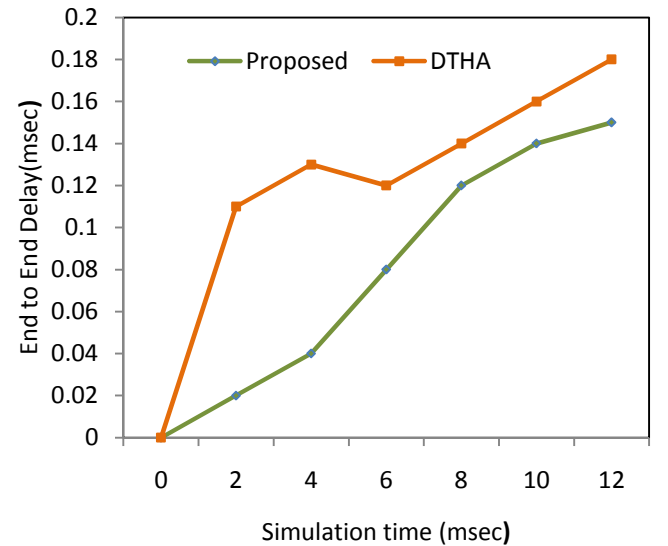


Fig. 2. End to End Delays(sec) of different nodes in the configured network

All these works of experimental studies were carried out with NS2 where the WiMax standards have verified through code and the parameters of the resultant graph. The area of operation in WiMax handoffs and base station analysis are made under coverage region of 1500X300 consisting of 25 mobile hosts and their buffer ranging parameters with respect to packet size sent between those mobile base stations. Thus all the values which are necessary for the study is collected and simulation is made according to the outcome for the experiment and accuracy in performance evolution is given evidently.

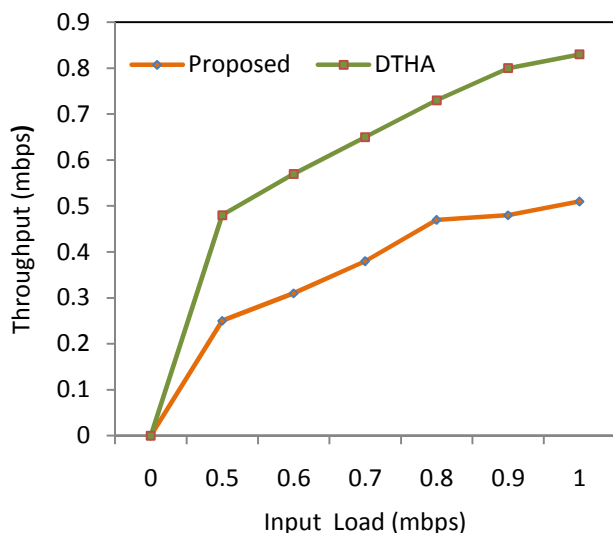


Fig. 3. Throughput of different nodes in the configured network

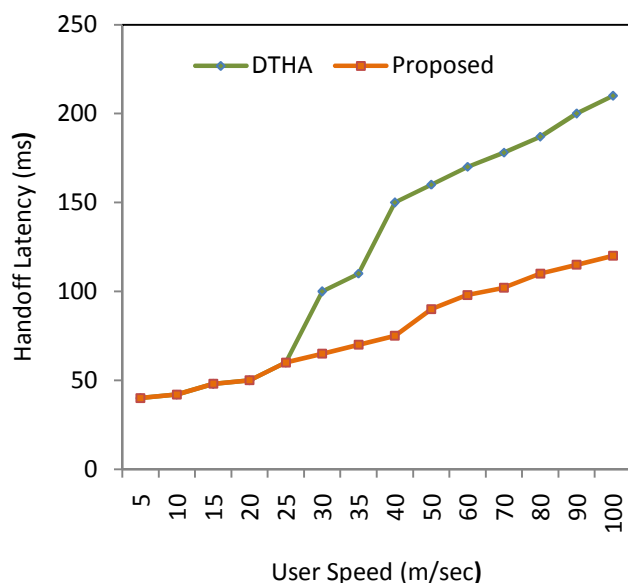


Fig. 4.: Handoff Latency

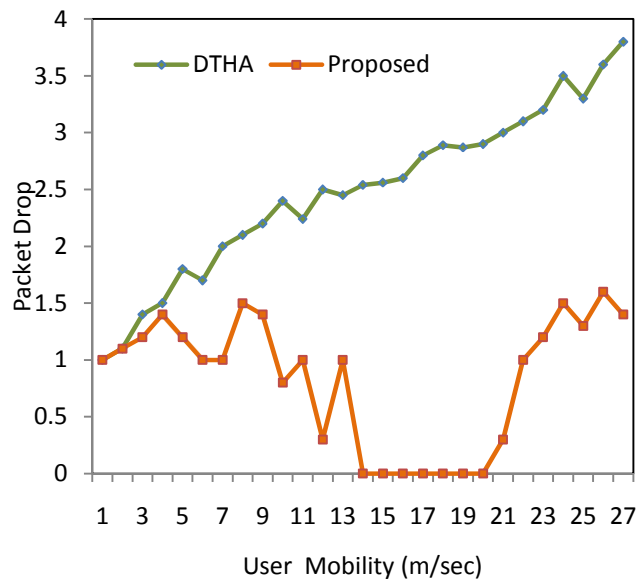


Fig. 5. Packet drop in communication network

6. CONCLUSION AND FUTURE WORK

In this paper we have constructed a handoff system that delivers elevated liveliness of bandwidth across the transmission medium with efforts from 802.16m handoff and dynamic buffering mechanism. Hence we can achieve Enhanced Quality of service and evaluated “Traffic Pacified Handoff algorithm” for the failure recovery process. This suggests that the proposed model and utilization of optimization routing may be useful in terms of dynamically altering buffer ranging and handoff thereby increasing profitability of increased throughput, retransmission that reduced overall round trip time delay and dynamically changing buffer size based on the bandwidth according to specific Quality of Service constraints.

REFERENCES

- [1] C. S. IEEE, LAN MAN Standards Committee of the, the IEEE Microwave Theory, and T. Society. IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1. IEEE Std 802.16e- 2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004), May 2006.
- [2] WiMAX and the IEEE 802.16m Air Interface Standard - April 2010
- [3] WiMAX: Broadband Wireless Access By Xiaole Song September 24, 2004
- [4] Jahangir khan, Ali Abbas, Khisro khan “Cellular Handover approaches in 2.5G to 5G Technology” Volume 21- No.2, May 2011

-
- [5] Ekici, E. "Optimal Two-Tier Cellular Network Design. BS Dissertation, Bogazici University, Turkey.2009.
- [6] Stallings, W. Wireless Communications & Networks. 2 nd Ed. USA, Prentice Hall. May 2007.
- [7] Jitendra R. Raol (2009). Multi-Sensor Data Fusion: Theory and Practice. CRC Press. ISBN 1-4398-0003-0.
- [8] Wei Chongyu, Liyong and Yang Wenlin, "Analysis on the RF interference in GSM/CDMA 1X dual-mode terminals" IEEE International Conference on Microwave and Millimeter Wave Technology, 2008. ICMMT 2008.
- [9] Hardeep Kaur and M L Singh. "Bit Error Rate Evaluation of IEEE 802.16 (WiMAX) in OFDM System" International Journal of Computer Applications 40(12):10-13, February 2012
- [10] Hyeong-Sook Park Sugrim, S., Spasojevic, P., Youn-Ok Park, "Noise Power and SNR Estimation Based on the Preamble in Tri-Sector OFDM Systems" Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd Date of Conference: 15-18 May 2011
- [11] Seon Ae Kim, Dong Geon An, Heung-Gyoon Ryu and Jin-up Kim, "Efficient SNR estimation in OFDM system", Radio and Wireless Symposium (RWS), 2011 IEEE Page(s): 182 – 185
- [12] N. Al-Rousan, O. Altrad, and Lj. Trajkovic, "Dual-trigger handover algorithm for WiMAX technology," OPNETWORK 2011, Washington, DC, Aug. 2011.
- [13] Sachin Lal Shrestha, Nah-Oak Songt, and Song Chong, "Seamless Realtime Traffic Handover Policy for IEEE 802.16m Mobile WiMAX", 43rd Annual Conference on Information Sciences and Systems, 2009.
- [14] Ioannis Papapanagiotou, Dimitris Toumpakaris, Jungwon Lee and Michael Devetsikioti, " A Survey on Next Generation Mobile WiMAX Networks: Objectives, Features and Technical Challenges", IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 11, NO. 4, 2009.
- [15] J. Mäkelä and K. Pentikousis, "Trigger Management Mechanisms", IEEE International Symposium on Wireless Pervasive Computing, San Juan, Puerto Rico, Feb 2007.
- [16] Mikko Majanen, Pekka H. J. Perälä, Thomas Casey, Jari Nurmi and Nenad Veselinovic "Mobile WiMAX Handover Performance Evaluation", Fifth International Conference on Networking and Services, 2009.

Enhancing Interaction Study on the Cloud Computing

Mohit Bhansali¹, Praveen Kumar², Seema Rawat³

¹Student, ²Assistant Professor, ³Assistant Professor

¹mohitbhansali@outlook.com, ²pkumar3@amity.edu, ³srawat1@amity.edu

Department of Computer Science and Engineering, Amity University, Noida, India

Abstract: Cloud Computing is evolving as a key technology for sharing resources. It is a generic term that involves high performance computing and storage infrastructure over the Internet. It is a new paradigm based on a pay-as-you-go approach. This paper introduces brief survey of cloud computing i.e. its architecture, platforms, explores the benefits, discuss the various elements of cloud computing issues i.e. security, privacy, reliability, open standard etc. and its application areas. It also covers the key technologies in Cloud Computing and Cloud Storage, different types of Cloud Services and describes the advantages and challenges of Cloud Storage. Clear insights into Cloud Computing will help the development and adoption of this evolving technology both for academe and industry.

Keywords: Cloud computing, Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Storage as a Service (StaaS).

1. INTRODUCTION

Cloud computing is the next generation in computation. Cloud computing is a metaphor for the internet. Cloud computing is a form of computing in which all applications, information and resource are managed in a virtual environment. Cloud computing involves virtual hosted environments allowing users to connect to the services being hosted over the internet. It is Internet ("cloud") based development and use of computer technology ("computing"). It is a style of computing in which IT-related capabilities are provided "as a service", allowing users to access technology-enabled services from the Internet (i.e., the Cloud) without knowledge of, expertise with, or control over the technology infrastructure that supports them. It implies Service Oriented Architecture (SOA). SOA is a way of reorganizing a portfolio of previous software applications and support infrastructure into an interconnected set of services, each accessible through standard interfaces and messaging protocols. Once all the elements of enterprise architecture are in place, existing and future applications can access these services as necessary without the need of convoluted point-to-point solutions based on inscrutable proprietary protocols. This architectural approach is particularly applicable when multiple applications running on varied technologies and platforms need to communicate

with each other. In this way, enterprises can mix and match services to perform business transactions with minimal programming effort.

A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. There is no need to have a server or any software to use it. All what is required is just an internet connection and e-mails are just one click away. The server and email management software is all on the cloud (i.e. internet) and is totally managed by the cloud service provider Yahoo, Google, AOL etc. The consumer gets to use the software alone and enjoy the benefits.

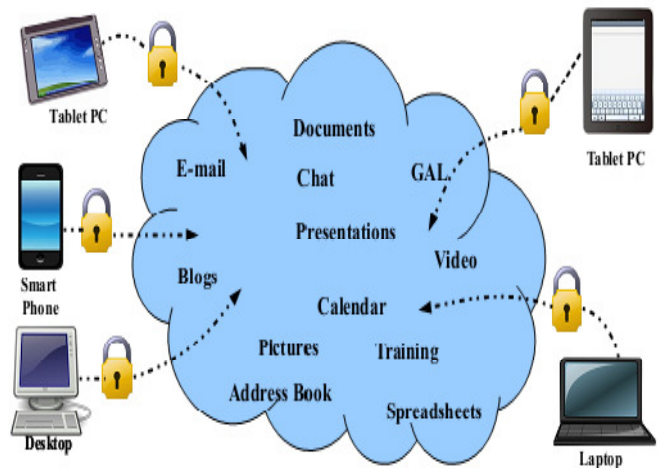


Fig. 1. Cloud Computing

2. WHY CLOUD COMPUTING

There are a lot of benefits of cloud computing over traditional computing. The major cloud providers such as Google, Microsoft and Amazon have built and are working on building the world's largest data centres across the United States and elsewhere. Each data centre includes hundreds of thousands of computer server, cooling equipment and substation power transformers. For example, consider Microsoft's data centre in Quincy, Washington. It has 43,600 square meters of space and uses 4.8 kilometres of chiller piping, 965 kilometres of electric wire, 92,900 square

meters of drywall and 1.5 metric tons of backup batteries. The company does not release the number of servers at this site; however it says that the data centre consumes 48 megawatts which is enough to power 40,000 homes [11]. As another example, the National Security Agency is planning to build a massive data centre at Fort Williams in Utah which is expected to consume over 70 megawatts electricity.

By the way, the cloud computing technology is environment friendly as by replacing the hardware with cloud computing systems reduces energy costs as well as reduces CO₂ emissions, it has business benefits because businesses can directly acquire the benefits of the huge infrastructure without having to implement and administer it directly, it can reduce implementation and maintenance costs having low initial cost, it has ease of backup system with compared to backing up all thick client PCs, it has mobility of information which easily used globally, it provides IT resources immediately and enables scalability according to needs of user of customer which is especially useful during peak times of the year when there is a need for additional resources that are not needed in other parts of the year. Users of cloud services have just operative expenses and capital expenses are minimized as much as possible. Usage of cloud computing services can foster innovation because there are no huge upfront costs for test and development environments.

Cloud computing will boost new markets which are already present in other business fields. This include cloud brokers which will be able to sell and buy resources like brokers do today on stock markets. Cloud computing also makes possible parallel batch processing which allows users to analyse terabytes of data for small periods of time and small costs, business analytics that can use the vast amount of computer resources for huge data warehousing.

Cloud computing provides the most reliable and secure data storage centre. Users do not have to worry about data loss, virus attach and other problems. The 'cloud' manages information by the professional team. Besides, it is easy and convenient to use. It is not necessary to download software and data or to upgrade dynamically in the 'cloud' side. We can access cloud services anytime and anywhere only with a computer connecting to the Internet. The benefits of deploying applications using cloud computing include reducing run time and response time, minimizing the risk of deploying physical infrastructure, lowering the cost of entry and increasing the pace of innovation.

3. HISTORY

The underlying concept of cloud computing was introduced way back in 1960s by John McCarthy. His opinion was that "computation may someday be organized as a public utility [1]." Also the characteristics of cloud computing were

explored for the first time in 1966 by Douglas Parkhill in his book, *The Challenge of the Computer Utility* [1]. The history of the term *cloud* is from the telecommunications world, where telecom companies started offering Virtual Private Network (VPN) services with comparable quality of service at a much lower cost. Initially before VPN, they provide dedicated point-to-point data circuits which was a wastage of bandwidth. But by using VPN services, they can switch traffic to balance utilization of the overall network. Cloud computing now extends this to cover servers and network infrastructure.

Many players in the industry have jumped into cloud computing and implemented it. Amazon has played a key role and launched the Amazon Web Services (AWS) in 2006. Also Google and IBM have started research projects in cloud computing. Eucalyptus became the first open source platform for deploying private clouds.

4. CHARACTERISTICS OF CLOUD COMPUTING

- In cloud computing, users access the data, applications or any other services with the help of a browser regardless of the device used and the user's location. The infrastructure which is generally provided by a third-party is accessed with the help of internet. Cost is reduced to a significant level as the infrastructure is provided by a third-party and need not be acquired for occasional intensive computing tasks.
- Less IT skills are required for implementation.
- Reliable service can be obtained by the use of multiple sites which is suitable for business continuity [1] and disaster recovery [1]. However, sometimes many cloud computing services have suffered outages and in such times its users can hardly do anything.
- Sharing of resources and costs amongst a large collection of users allow efficient utilization of the infrastructure.
- Maintenance is easier in case of cloud computing applications as they need not be installed on each user's computer.
- Pay per use facility allows measuring the usage of application per client on regular bases.
- Performance can be monitored and thus it is scalable.
- Security can be good as or better than traditional systems because providers are able to devote resources to solving security issues that many customers cannot afford. However, security still remains an important concern when the data is quite confidential. This delays adoption of cloud computing to some extent.

5. CLOUD COMPUTING – DEPLOYMENT AND DELIVERY MODELS

Cloud Computing can be classified and deployed in a number of ways e.g. as *public*, *private* or *hybrid* clouds.

1) *Public Clouds*: also known as External Clouds, are cloud services provided by third parties and hosted and managed by the service providers. The cloud providers assume the responsibilities of installation, management, provisioning and maintenance. The variety of Clouds provides a much greater level of efficiency of pooling of resources. The customers access and consume the services and IT resources. Consumers are charged only for the resources and services they are following a pay-as-you-go approach. Lack of appropriate security, reliability and regulatory compliance is often a major issue here. Amazon.com is one of the largest public cloud providers.

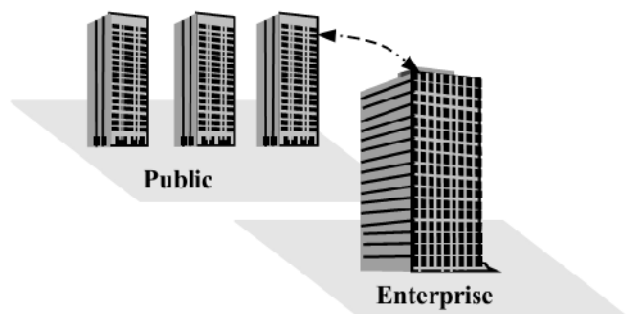


Fig. 2: Public Cloud

2) *Private Clouds*: also known as Internal Clouds, are proprietary networks, often data centres, residing within the enterprise for the exclusive use of the organization or for a known group of consumers. A local or private network infrastructure is employed. In this case, the enterprise is in charge of setting up and maintaining the cloud and thus the enterprise can take better control of all aspects of the provision and functioning. The added advantage is in terms of better control of security, more effective regulatory compliance and improved quality of service. For mission critical processes and for location of sensitive data, this type cloud infrastructure is much more privacy than a Public Cloud would.

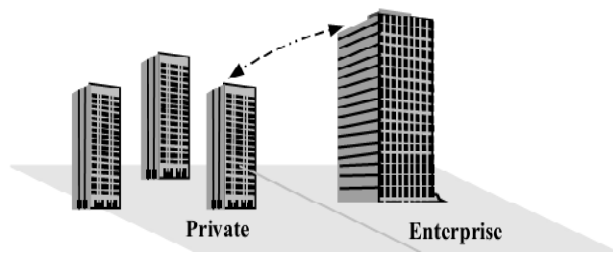


Fig. 3: Private Cloud

Private Clouds are, generally, Clouds that reside within the organization, however, private clouds, outside the organization, are also becoming a possibility, where the resources inside such a Cloud are available only to the organization concerned and totally invisible to others.

When a service provider uses public cloud resources to create their private cloud, the result is a *virtual private cloud*. A Community Cloud is a semi-private cloud that is used by the defined group of tenants with shared backgrounds and requirements [17]. This, then, becomes a private cloud for this community, where the management responsibility is shared amongst the members of the community.

3) *Hybrid Clouds*: are a combination of private and public clouds. In this case, the management responsibilities are split between the enterprise and the public cloud providers, which can often become an issue of concern. For mission critical processes, this type of cloud infrastructure can also be highly effective because of enhanced control and management by the enterprise itself. For example, the organization can keep the sensitive data within the private cloud and the rest in the public cloud.

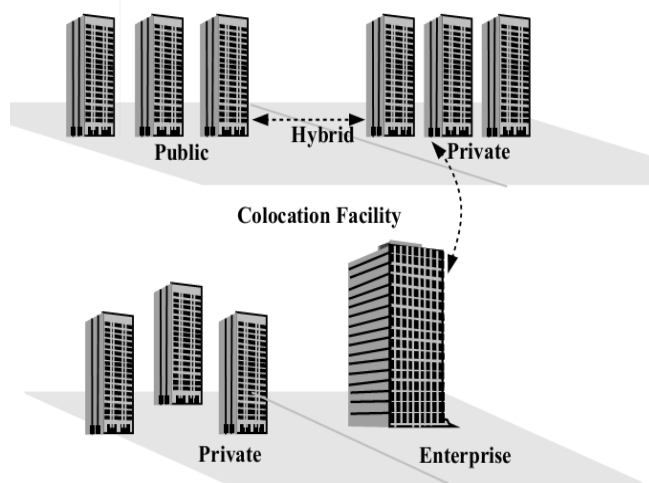


Fig. 4: Hybrid Cloud

The Cloud model generally consists of three varieties of architecture which refer to and provide three types of generic services, namely: *Software Services*, *Platform Services*, and *Infrastructure Services*. These are generally abbreviated as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS), respectively.

Presenting the model as pyramid, the Software Services will be at the top and the Infrastructure Services will be the bottom of the pyramid. Based on this anatomy, the Cloud

Services are often defined, as in the following sections and as shown in Figure below.

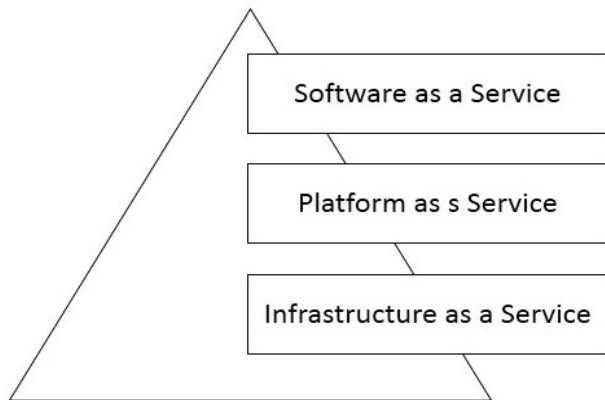


Fig. 5. A 3-Layer Model of Cloud Computing

A. Software as a Service

This refers to prebuilt and vertically integrated applications are delivered to and purchased by customers as *services*. Here, customers are looking to purchase functionality. Applications are normally designed for ease of use based on proven business models. This may be regarded as a user level layer and it can be further classified into separate layers: 1) Services (which are often standalone applications e.g. a billing service); and 2) Applications (which are often units of functionalities). SaaS is a very broad market where services can be anything from Web-based email to inventory control, even in some cases outline banking services, as well as database processing. Gmail, Hotmail, Quicken Online, IBM®Websphere, Boomi and Salesforce are some of the well-known SaaS products and providers.

B. Platform as a Service

This layer refers to software and product development tools (e.g. application servers, database servers, portal servers, middleware, etc.) which clients purchase so they can build and deploy their own applications, thus providing a much increased flexibility and control to the consumer. However, there may sometimes be a certain amount of dependence upon the infrastructure and platform providers. The services, here, are intended to support the ‘software services’ top layer of the pyramid. The customers are looking to buy time and cost savings in deploying their applications. Typical offerings include runtime environment for application code, Cloud Services, Computer power, storage, and networking infrastructure. This level of services may be regarded as a *developer level* layer. The pricing structure is often along the lines of: Compute usage per hour; data transfer per GB; IO requests per million; storage per GB; data storage requests per thousand. All charges are per each billing period. Google

App Engine, Heroku, Mosso and Engine Yard are examples of PaaS products and providers.

C. Infrastructure as a Service

This layer is essentially hardware (e.g. visualized servers, storage, network devices, etc.) and hardware services to enable Cloud Platforms and Applications to operate. These services support the ‘software services’ top layer of the pyramid. Customers get full control over server infrastructure and that sometimes comes with a price premium. Here, customers are looking to buy ‘computing’, without making upfront investment. Since, the infrastructure is offered on a pay-for-what-you-see basis, it is sometimes referred to as *utility computing*, as there is similarity with the provision and use of services such as electricity and gas. The pricing structure is often similar to the provision for PaaS. Amazon EC2, IBM BlueHouse, VMWare, GoGrid, RightScale and Linode are some of the IaaS products and providers.

D. Storage as a Service

Commonly known as Storage as a Service (StaaS), it facilitates cloud applications to scale beyond their limited servers. StaaS allows users to store their data at remote disks and access them anytime from any place. Cloud storage systems are expected to meet several rigorous requirements for maintaining users’ data and information, including high availability, reliability, performance, replication and data consistency; but because of the conflicting nature of these requirements, no one system implements all of them together.

E. Other Provision as Services

The dividing line between the three layers as shown in Figure 1 is not clear and, in fact there is a considerable amount of overlap. For example, a software system may be considered as part of a software platform; similarly, an IS platform may be considered as part of IS infrastructure. It is for this reason that researchers have also discussed combined models such as: SaaS & PaaS; SaaS & IaaS; IaaS & PaaS; and even SaaS & PaaS & IaaS. Numerous other categories have also been suggested in recent years e.g.:

- Database-as-a-Service
- Security-as-a-Service
- Communication-as-a-Service
- Management/ Governance-as-a-Service
- Integration-as-a-Service
- Testing-as-a-Service
- Business Process-as-a-Service

In this respect, any provision that is available and that provides support in some sense to the consumers is regarded as a ‘service’. For an enterprise, it is not enough to have services available in the Cloud. There is, often, also a requirement of expertise available to help the enterprise to interface, sequence, and integrate the services with what already exists. So, Integration-as-a-Service or ‘Solution-as-a-Service’ can be particularly important service.

Figure shows the various cloud computing services with their examples.

SaaS	PaaS	IaaS	StaaS
Software as a Service	Platform as a Service	Infrastructure as a service	Storage as a Service
1.Communication (email) 2.Collaboration 3.Productivity tools 4.ERP	1.Application Development 2.Security Services 3.Database Management	1.Servers 2.Network 3.Storage 4.Management 5.Reporting	1.Primary 2.Backup 3.Archive 4.DR
Examples	Examples	Examples	Examples
SalesForce.com NetSuite Oracle IBM Google Apps	GAE Microsoft’s Azure Amazon EC2	GoGrid Flexiscale Joyent	Amazon S3 Nirvanix

All the above mentioned services are pay per use, which makes cloud computing an attractive option for those organization which cannot afford buying, installing and maintain the required services.

6. CLOUD COMPUTING ISSUES

Cloud computing has grown from being a promising business concept to one of the fastest growing segments of the IT industry. Now, recession-hit companies are increasingly realizing that simply by tapping into the cloud they can gain fast access to best-of-breed business applications or drastically boost their infrastructure resources, all at negligible cost. But as more and more information on individuals and companies is placed in the cloud, concerns are beginning to grow about just how safe an environment it is.

A. Security

Where is your data more secure, on your local hard driver or on high security servers in the cloud? Some argue that customer data is more secure when managed internally, while other argue that cloud providers have a strong incentive to maintain trust and as such employ a higher level of security. However, in the cloud, your data will be distributed over these individual computers regardless of

where your base repository of data is ultimately stored. Industrious hackers can invade virtually any server, and there are the statistics that show that one-third of breaches result from stolen or lost laptops and other devices and from employees’ accidentally exposing data on the Internet, with nearly 16 percent due to insider theft [2].

B. Privacy

Different from the traditional computing model, cloud computing utilizes the virtual computing technology, users’ personal data may be scattered in various virtual data centre rather than stay in the same physical location, even across the national borders, at this time, data privacy protection will face the controversy of different legal systems. On the other hand, users may leak hidden information when they accessing cloud computing services. Attackers can analyse the critical task depend on the computing task submitted by the users [3].

C. Reliability

Servers in the cloud have the same problems as your own resident servers. The cloud servers also experience downtimes and slowdowns, the difference is that users have a higher dependent on cloud service provider (CSP) in the model of cloud computing. There is a big difference in the CSP’s service model, once you select a particular CSP, you may be locked-in, thus bring a potential business secure risk.

D. Legal Issues

Regardless of efforts to bring into line the lawful situation, as of 2009, supplier such as Amazon Web Services provide to major markets by developing restricted road and rail network and letting users to choose “availability zones” [4]. On the other hand, worries stick with safety measures and confidentiality from individual all the way through legislative levels.

E. Open Standard

Open standards are critical to the growth of cloud computing. Most cloud providers expose APIs which are typically well-documented but also unique to their implementation and thus not interoperable. Some vendors have adopted others’ APIs [5] and there are a number of open standards under development, including the OGF’s Open Cloud Computing Interface. The Open Cloud Consortium (OCC) [6] is working to develop consensus on early cloud computing standards and practices.

F. Compliance

Numerous regulations pertain to the storage and use of data require regular reporting and audit trails, cloud providers must enable their customers to comply appropriately with

these regulations. Managing Compliance and Security for Cloud Computing, provides insight on how a top-down view of all IT resources within a cloud-based location can deliver a stronger management and enforcement a compliance policies. In addition to the requirements to which customers are subject, the data centres maintained by cloud providers may also be subject to compliance requirements [7].

G. Freedom

Cloud computing does not allow users to physically possess the storage of the data, leaving the data storage and control in the hands of cloud providers. Customers will contend that this is pretty fundamental and affords them the ability to retain their own copies of data in a form that retains their freedom of choice and protects them against certain issues out of their control whilst realizing the tremendous benefits cloud computing can bring [8].

H. Long-term Viability

You should be sure that the data you put into the cloud will never become invalid even your cloud computing provider go broke or get acquired and swallowed up by a larger company. “Ask potential providers how you would get your data back and if it would be in a format that you could import into a replacement application,” Gartner says [9].

7. APPLICATIONS AND CHALLENGES OF CLOUD COMPUTING

A. Applications

The system has proven to be favoured by the users over time and we now have many other players in the arena. A famous one in Google Apps, which offers email, calendar, document editing and more in the cloud. Even Microsoft, which arguably benefited most from local computing, is increasing its focus on cloud computing services now [10]. The applications of cloud computing are practically limitless. With the right middleware a cloud computing system could execute all the programs a normal computer could run. Potentially, everything from generic word processing software to customized computer programs designed for a specific company could work on a cloud computing system.

B. Challenges

The cloud computing is an emerging technology but there are several challenges which needs to be looked upon. These are described briefly below.

1) *Regulatory Compliance*: When outsourcing to a provider, customers are responsible for the security and integrity of their own data, even when it is held by a third party provider.

2) *Dependency*: It is only possible to use applications or services that the provider is willing to offer.

3) *Data Location & Privacy Restrictions*: US & EU have different privacy standards, subject to different laws.

4) *Recovery*: Data segmentation makes back-ups more difficult.

5) *Data Storage*: Cloud computing does not allow users to physically store of their data, so data storage is done by the provider.

6) *Data security and privacy protection*: The security of user data is considered to be the security problem of computing platforms, security problem of computing platform is an important issue of cloud computing. Cloud computing infrastructure with a multi-tenant properties, manufactures generally cannot guarantee that the data of two different users to achieve physical separation.

7) *Data access and storage model*: There are simple memory model or a simple hierarchical model which based on binary object. It has brought significant flexibility, it also increase the burden to the application logic explains the relationship between different data elements.

8) *Lack of standards and vendor locking*: Most vendors have defined standards based mechanism e.g. HTTP, REST (REpresentation State Transfer), SOAP (Simple Object Access Protocol) etc. to access and use its services. However, the standard of development services in cloud computing is just rising and now the lack of function of write once and run everywhere.

9) *Services Interoperability*: Cloud computing doesn't have enough support for the interoperability of services, this effects the cross-platform services.

8. ADVANTAGES OF CLOUD COMPUTING

a. Easy Management

The maintenance of the infrastructure, be it hardware or software is simplified, thus, less headaches for the IT team. Also applications that are quite storage extensive are easier to use in the cloud environment compared to the same when used by the organization by its own. Also at the user level, what you mostly need is a simple browser with internet connectivity.

b. Cost Reduction

The main advantage for SMEs lies here. Cloud computing drastically reduces the IT spending for SMEs. Costly systems need not be required for occasional use of intensive

computing resources. Also the main power required for such systems is not required. Even simple applications like email can be set up and most free through applications like Google Apps. Also as most of the time such providers are quite reliable in terms of availability, it is clear winner.

c. Uninterrupted Services

Lower outages are provided by cloud computing services, this providing uninterrupted services to the user. However, some occurrences of outages have occurred in the past, like the Gmail outage in 2009. Also other cloud vendors like EC2 have failed at some point of time, but however, they are much more dependable compared to the infrastructure installed on the organization.

d. Disaster Management

In case of disasters, an offsite backup is always helpful. Keeping crucial data backed up using cloud storage services is the need of the hour for most of the organization. Also cloud storage services not only keep your data off site, but they also ensure that they have systems in place for disaster recovery.

e. Green Computing

Harmful emissions due to extensive use of systems in organizations, electronic waste generated as the time passes and energy consumption is the main disadvantage of the present day computing systems. This can be reduced to some extent by using cloud computing services. This leads to environment preserving. Also the e-waste is generated to minimum extent.

9. CONCLUSION

This paper presents, in some detail, the deployment approaches, the benefits, the issues and challenges. The aim of this work is to provide some general information for enterprises wishing to integrate their existing IT processes and system with Cloud infrastructure available outside their organization. It is predicted that this technology brings for us an infinite capability of computing, fast microprocessor, huge memory, high-speed network, reliable system architecture etc. by solving the existing issues and challenges and we will enter a new era of next generation computing through cloud computing technology.

REFERENCES

- [1] Elinor Mills, January 27, 2009. "Cloud computing security forecast: clear skies".
- [2] Randolph Barr, Qualys Inc, "How to gain comfort in losing control to the cloud".
- [3] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, [2005] "Live migration of virtual machines" In Proc. of NSDI'05, pages 273-286, Berkeley CA, USA, 2005. USENIX Association.
- [4] Eucalyptus Completes Amazon Web Services Specs with Latest Release.
- [5] Open Cloud Consortium.org.
- [6] July 27, 2009. Available from <http://fx.caixun.com/>.
- [7] Jack Schofield. Wednesday 17 June 2009 22.00 BST, <http://www.guardian.co.uk/technology/2009/jun/17/cloud-computingjack-schofield>.
- [8] Gartner. "Seven cloud-computing security risks". <http://www.infoworld.com> July 02, 2008.
- [9] Weinman, J.; "The future of Cloud Computing," Technologies Beyond 2020 (TTM), 2011 IEEE Technology Time Machine Symposium on, vol., no., pp.1-2, 1-3 June 2011
- [10] Katz, R.H.; "Tech Titans Building Boom," Spectrum, IEEE, vol.46, no.2, pp.40-54, Feb. 2009
- [11] Kalagiakos, P.; Karampelas, P.; , "Cloud Computing learning," Application of Information and Communication Technologies (AICT), 2011 5th International Conference on , vol., no., pp.1-4, 12-14 Oct. 2011
- [12] Jadeja, Y.; Modi, K.; "Cloud computing - concepts, architecture and challenges," Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on, vol., no., pp.877-880, 21-22 March 2012
- [13] Mollah, M.B.; Islam, K.R.; Islam, S.S.; "Next generation of computing through cloud computing technology," Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on, vol., no., pp.1-6, April 29 2012-May 2 2012
- [14] Mahmood, Z.; "Cloud Computing: Characteristics and Deployment Approaches," Computer and Information Technology (CIT), 2011 IEEE 11th International Conference on, vol., no., pp.121-126, Aug. 31 2011-Sept. 2 2011
- [15] Rajan, S.; Jairath, A.; "Cloud Computing: The Fifth Generation of Computing," Communication Systems and Network Technologies (CSNT), 2011 International Conference on, vol., no., pp.665-667, 3-5 June 2011
- [16] Caroline Kvitka, Clouds Bring Agility to the Enterprise, <http://www.oracle.com/technology/oramag/oracle/10-mar/o20interview.html>

Overview of IPv6 and its Implementation

Nistha Rai¹, Kanak Priya², Neha Kumari³

M.Tech (TSE), AITTM, Amity University, Noida
rai.nistha@gmail.com¹, kanak.priya90@gmail.com², nehakumari49@yahoo.com³

Abstract: Today, most electronic devices such as mobile phones, PCs, Internet telephones, etc use in homes and other places, rely on the internet technology for their various services. The internet connected devices use the internet protocol (IP) address to communicate over the network with each device assigned a unique IP address. This means that, for any device to communicate through the internet, it must be assigned an IP address. Hence, the tremendous growth rate in the number of internet connected devices and high dependence on the internet for human daily activities have caused the expected exhaustion of the long-time used IP addresses. Internet Protocol version 6 (IPv6) or IP Next Generation is the protocol that has been designed to replace the existing Internet Protocol version 4 (IPv4). In this paper we study about IPv6 why it was needed and the advantages of IPv6 over IPv4 and its implementation in India

Keywords: Internet Protocol, IPv6, IPv4

1. INTRODUCTION

The Network layer in TCP/IP protocol suite is currently IPv4. IPv4 provides host to host communication between the systems in the internet. It was developed in 1970, at the beginning of the internet, and has not been substantially changed since it was published in 1981. IPv4 is an unreliable and connectionless datagram protocol – a best-effort delivery service. The term best-effort means that IPv4 provides no error control or flow control (except for error detection on the header) to get a transmission through to its destination, but with no guarantees. Although IPv4 has proven to be robust, well designed, easily implemented and interoperable, the initial design does not anticipate the following;

- The exponential growth of the Internet and the impending exhaustion of the IPv4 address space.
- The ability of internet backbone routers to maintain large routing tables.
- The need for simpler and automatic configuration of IP addresses.
- The requirement of security at IP layer.
- The need for better support for real-time delivery of data also called quality of service (QOS) for applications like VOIP, VOD etc.

2. INTERNET PROTOCOL VERSION 6 (IPv6)

To overcome the deficiencies of IPv4, IPv6 (Internetworking Protocol, version 6), also known as IPng (Internetworking protocol, next generation) was proposed and now is a standard. It was developed by the IETF in the mid Nineties. In IPv6, the internet protocol was extensively modified to accommodate the unforeseen growth of the internet. IPv6 improves on the addressing capacities of IPv4 by using 128 bits for addressing instead of 32, thereby making available an almost infinite pool of IP addresses. The format and the length of the IP address was changed along with the packet format. Related Protocols, such as ICMP, were also modified. Other protocols in the network layer, such as ARP, RARP and IGMP, were either deleted or included in ICMPv6 protocol. Routing protocols such as RIP and QSPF were also slightly modified to accommodate these changes. In addition, IPv6 is supposed to be providing various enhancements with respect to security, routing, address auto configuration, mobility & QOS etc.

3. IMPORTANT FEATURES OF IPV6:

The following are the important features of IPv6 protocol, which may play an important role in the growth of Internet in the country due to its advance capabilities.

(i) New Header Format

The IPv6 header has a new format that is designed to keep header overhead to a minimum. The streamlined IPv6 header is more efficiently processed at intermediate routers with lower processing costs.

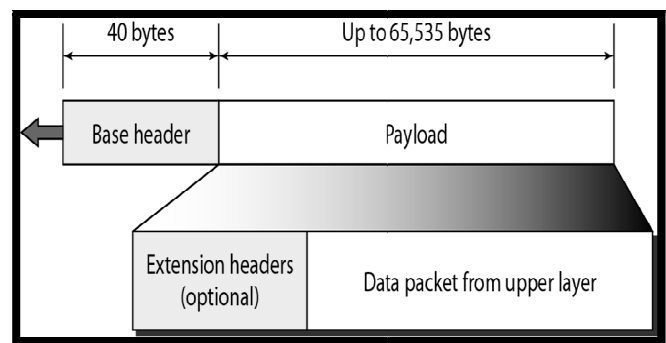


Fig. IPv6 datagram header diagram

(ii) Large Address Space

IPv6 has 128 bits (16 bytes) source and destination IP addresses. This enables it to accommodate 2^{128} hosts. Even though only a small number of IPv6 addresses are currently allocated for use by hosts, there are plenty of addresses available for future use. With a much larger number of available addresses, address conservation techniques, such as deployments of NAT is no longer necessary.

(iii) Efficient and Hierarchical Addressing and Routing Infrastructure

IPv6 global addresses used on the IPv6 portion of the Internet are designed to create an efficient, hierarchical, and summarisable routing infrastructure that is based on the common occurrence of levels of Internet service providers.

(iv) Stateless and stateful address configuration

IPv6 supports both stateful address configuration, such as address configuration in the presence of a DHCP server, and stateless address configuration (address configuration in the absence of a DHCP server). With stateless address configuration, hosts on a link automatically configure themselves with IPv6 addresses for the link (called link-local addresses) and with addresses derived from prefixes advertised by local routers. Even in the absence of a router, hosts on the same link can automatically configure themselves with link-local addresses and communicate without manual configuration.

(v) Built-in Security

Support for IPsec is an IPv6 protocol suite requirement. This requirement provides a standards-based solution for network security needs and promotes interoperability between different IPv6 implementations.

(vi) Support for QOS

New fields in the IPv6 header define how traffic is handled and identified. Traffic identification using a Flow Label field in the IPv6 header allows IPv6 routers to identify and provide special handling for packets belonging to particular packet flow between source and destination. Support for QOS can be achieved even when the packet payload is encrypted through IPsec.

(vii) Extensibility

IPv6 can easily be extended for new features by adding extension headers after the IPv6 header.

4. TRANSITION FROM IPV4 TO IPV6

Because of the huge number of systems on the internet, the transition from IPv4 to IPv6 cannot happen suddenly. It

takes a considerable amount of time before every system in the internet can move from IPv4 to IPv6. The transition must be smooth to prevent any problems between IPv4 to IPv6 systems. Three strategies have been devised by the IETF to help the transitions.

1. Dual Stack

The dual IP layer is an implementation of the TCP/IP suite of protocols that includes both an IPv4 Internet layer and an IPv6 Internet layer. This is the mechanism used by IPv6/IPv4 nodes so that communication with both IPv4 and IPv6 nodes can occur. A dual IP layer contains a single implementation of Host-to-Host layer protocols such as TCP and UDP. All upper layer protocols in a dual IP layer implementation can communicate over IPv4, IPv6, or IPv6 tunnelled in IPv4. It is recommended that all hosts, before migrating completely to version 6, have a dual stack of protocols. In other words, a station must run IPv4 and IPv6 simultaneously until all the internet uses IPv6. To determine which version to use when sending a packet to a destination, the source host queries the DNS. If the DNS returns an IPv4 address, the source host sends an IPv4 packet. If the DNS returns an IPv6 address, the source host sends an IPv6 packet.

2. Tunneling

Tunneling is the strategy when two computers using IPv6 wants to communicate with each other and the packet must pass through a region that uses IPv4. To pass through this region, the packet must have an IPv4 address. So the IPv6 packet is encapsulated in an IPv4 Packet when it enters the region, and it leaves its capsule when it exits the region. It seems as if the IPv6 packet goes through a tunnel at one end and emerges at the other end. To make it clear that the IPv4 packet is carrying an IPv6 packet as data, the protocol value is set to 41.

3. Header Translation Strategy

Header Translation is necessary when the majority of the internet has moved to IPv6 but some systems still use IPv4. The sender wants to use IPv6, but the receiver does not understand IPv6. Tunneling does not work in this situation because the packet must be in IPv4 format to be understood by the receiver. In this case, the header format must be totally changed through header translation. The header of the IPv6 packet is converted to an IPv4 header. Header translation uses the mapped address to translate an IPv6 address to an IPv4 address.

5. IPV6 DEPLOYMENT STATUS IN THE WORLD

1. China instituted a full adoption policy of IPv6 by creating the China Next Generation Internet (CNGI).

2. The European Commission has shown strong support for IPv6 with the creation of the EU IPv6 Task Force.
3. Moonv6 is an international project led by NAv6TF to execute deployment testing of IPv6 technology.
4. The South Korean public sector has been engaged in deploying IPv6 on national level by building a nation-wide IPv6 MPLS backbone.
5. The Japanese Government has taken on a program initiative called “u-Japan”. It is centred on empowering the Japanese end-user.

6. IPV6 DEPLOYMENT STATUS IN INDIA

1. Various issues on IPv6 were deliberated taken at different levels in DoT, TRAI and TEC.
2. Suitable policy framework by Govt. for smooth Transition.
3. Creation of IPv6 Task Force and working Groups.
4. Govt. departments are taking IP-based services from only IPv6 ready ISPs.

7. INDIAN GOVERNMENT FUTURE STRATEGIES

1. Substantial transition to IPv6 by 2020
2. IPv6 innovation centre is set up.
3. All the Govt websites are on dual stack from the end of 2012.
4. Adoption of IPv6 based pilot projects in Govt sectors.

5. IPv6 implementation workshops are created for states and central organizations.

8. CONCLUSION

Currently used IPv4 addresses are depleting. Due to increasing population and daily inventions of new technologies more IP addresses are needed. Solution is IPv6. There are 3 transition mechanisms for conversion of IPv4 into IPv6. Transition to IPv6 is not easy and will not happen overnight. Organizations have invested lot of money in building the IPv4 infrastructure over the years and replacement of that infrastructure is not feasible without the recovery of the investment. Therefore, IPv4 and IPv6 will co-exist for a long time to come. IPv6 Implementation is still in process in India and world.

9. ACKNOWLEDGEMENTS

It is a great pleasure to acknowledge my profound sense of gratitude to my project guide Mrs. Neha Arora, Assistant Professor, AITTM AUUP for her valuable and inspiring guidance, comments, suggestions and encouragement throughout this paper.

REFERENCES

- [1] Geoff Huston, APNIC, “IPv4 Address Depletion and Transition to IPv6”, Internet Protocol Journal, Vol. 10, No. 3, pp 18-28, 2007.
- [2] Geoff Huston, Telstra, “IPv4: How long do we have?”, Internet Protocol Journal, Vol. 6, No. 4, pp 2-15, 2003.
- [3] Gregory R. Schloz, Clint Evans, Jaime Flores, Mustafa Rahman, “Internet protocol version 6”, Internet Protocol Journal, Vol. 16 , Issue 3 , pp 197 – 204, March 2001. and services, Sophia Antipolis, France, Sept. 2002.

Synchronization Techniques for OFDM Systems

Pratima Manhas¹, Shaveta Thakral²

^{1,2}Deptt. of Electronics and Communications, MRIU, Faridabad
¹pratimamehak@gmail.com, ²Shwithakral167@gmail.com

Abstract: Orthogonal frequency division multiplexing (OFDM) is a multi-carrier transmission technique, which divides the available spectrum into many subcarriers, each one being modulated by a low data rate stream. Synchronization is of great importance for all digital communication systems. OFDM systems are very sensitive to both timing and carrier frequency offset, especially, when combined with other multi-access techniques such as FDMA, TDMA, and CDMA. Therefore, synchronization is extremely crucial to the OFDM systems.

Keywords: OFDM, Synchronization, multi- carrier, subcarriers

1. INTRODUCTION

OFDM stands for 'Orthogonal Frequencies Division Multiplexing'. Instead of modulating a single carrier as is the case with FM or AM, the idea to utilize a number of carriers, spread regularly over a frequency band, in such a way that each of them is modulated with just a small portion of the available energy. This can be done in a flexible way so that the available bandwidth is utilized to maximal efficiency. It is also a 'broadband' technique in that the modulated signal fills out a large bandwidth almost by definition

The first OFDM schemes presented in 1966. It distributes the data over a large number of carriers that are spaced apart at precise frequencies. This spacing provides the "orthogonality" in this technique which prevents the demodulators from seeing frequencies other than their own. The principal reason of this increasing interest is due to its capability to provide high-speed data rate transmissions with low complexity and to counteract the intersymbol interference (ISI) introduced by dispersive channels. For this reason OFDM modulation has been adopted by several digital wireline and wireless communication standards, such as the European digital audio and video broadcasting standards, as well as local area networks.

Due to the high data rate transmission and the ability to against frequency selective fading, orthogonal frequency division multiplexing (OFDM) is a promising technique in the current broadband wireless communication system.

Orthogonal frequency division multiplexing (OFDM) technology is to split a high-rate data stream into a number of lower rate streams that are transmitted simultaneously over a number of subcarrier. Because the symbol duration

increases for the lower rate parallel subcarrier, the relative amount of dispersion in time caused by multipath delay spread is decreased. In OFDM systems, the spectrum of individual subcarrier is overlapped with minimum frequency spacing, which is carefully designed so that each subcarrier is orthogonal to the other subcarriers. The bandwidth efficiency of OFDM is another advantage.

Basic idea is to use a large number of parallel narrow-band subcarriers instead of a single wide-band carrier to transport information

OFDM can be viewed as either a modulation technique or a multiplex technique.

- Modulation technique

-Viewed by the relation between input and output signals

- Multiplex technique

-Viewed by the output signal which is the linear sum of the modulated signals

2. OFDM MODEL

If the FDM system above had been able to use a set of subcarriers that were orthogonal to each other, a higher level of spectral efficiency could have been achieved. The guard bands that were necessary to allow individual demodulation of subcarriers in an FDM system would no longer be necessary. The use of orthogonal subcarriers would allow the subcarriers' spectra to overlap, thus increasing the spectral efficiency. As long as orthogonality is maintained, it is still possible to recover the individual subcarriers' signals despite their overlapping spectrums.

If the dot product of two deterministic signals is equal to zero, these signals are said to be orthogonal to each other. Orthogonality can also be viewed from the standpoint of stochastic processes. If two random processes are uncorrelated, then they are orthogonal. Given the random nature of signals in a communications system, this probabilistic view of orthogonality provides an intuitive understanding of the implications of orthogonality in OFDM. OFDM is implemented in practice using the discrete Fourier transform (DFT). The sinusoids of the DFT form an

orthogonal basis set, and a signal in the vector space of the DFT can be represented as a linear combination of the orthogonal sinusoids. One view of the DFT is that the transform essentially correlates its input signal with each of the sinusoidal basis functions. Mathematically

DFT:

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi nk/N)$$

IDFT:

$$x(k) = (1/N) \sum_{n=0}^{N-1} X(n) \exp(j2\pi nk/N)$$

Example:

For $N=2$ if we calculate the basis, it comes out to be $[1, 1]$ and $[1, -1]$

Here we can see that this is orthogonal vectors. And one of this is a DC sinusoidal component while another is with some frequency. So we exploit this orthogonal sinusoidal basis property in OFDM

If the input signal has some energy at a certain frequency, there will be a peak in the correlation of the input signal and the basis sinusoid that is at that corresponding frequency. This transform is used at the OFDM transmitter to map an input signal onto a set of orthogonal subcarriers, i.e., the orthogonal basis functions of the DFT. Similarly the transform is used again at the OFDM receiver to process the received subcarriers. The signals from the subcarriers are then combined to form an estimate of the source signal from the transmitter. The orthogonal and uncorrelated nature of the subcarriers is exploited in OFDM with powerful results. Since the basis functions of the DFT are uncorrelated, the correlation performed in the DFT for a given subcarrier only sees energy for that corresponding subcarrier. The energy from other subcarriers does not contribute because it is uncorrelated. This separation of signal energy is the reason that the OFDM subcarriers' spectrums can overlap without causing interference.

The idea behind the analog implementation of OFDM can be extended to the digital domain by using the discrete Fourier Transform (DFT) and its counterpart, the inverse discrete Fourier Transform (IDFT). These mathematical operations are widely used for transforming data between the time-domain and frequency-domain. These transforms are interesting from the OFDM perspective because they can be viewed as mapping data onto orthogonal subcarriers. For example, the IDFT is used to take in frequency-domain data and convert it to time-domain data. In order to perform that operation, the IDFT correlates the frequency-domain input data with its orthogonal basis functions, which are sinusoids

at certain frequencies. This correlation is equivalent to mapping the input data onto the sinusoidal basis functions.

In practice, OFDM systems are implemented using a combination of fast Fourier Transform (FFT) and inverse fast Fourier Transform (IFFT) blocks that are mathematically equivalent versions of the DFT and IDFT, respectively, but more efficient to implement. An OFDM system treats the source symbols (e.g., the QPSK or QAM symbols that would be present in a single carrier system) at the transmitter as though they are in the frequency-domain. These symbols are used as the inputs to an IFFT block that brings the signal into the time-domain. The IFFT takes in N symbols at a time where N is the number of subcarriers in the system. Each of these N input symbols has a symbol period of T seconds. The basis functions for an IFFT are N orthogonal sinusoids. These sinusoids each have a different frequency and the lowest frequency is DC. Each input symbol acts like a complex weight for the corresponding sinusoidal basis function. Since the input symbols are complex, the value of the symbol determines both the amplitude and phase of the sinusoid for that subcarrier. The IFFT output is the summation of all N sinusoids. Thus, the IFFT block provides a simple way to modulate data onto N orthogonal subcarriers. The block of N output samples from the IFFT make up a single OFDM symbol. The length of the OFDM symbol is NT where T is the IFFT input symbol period (shown in fig1).

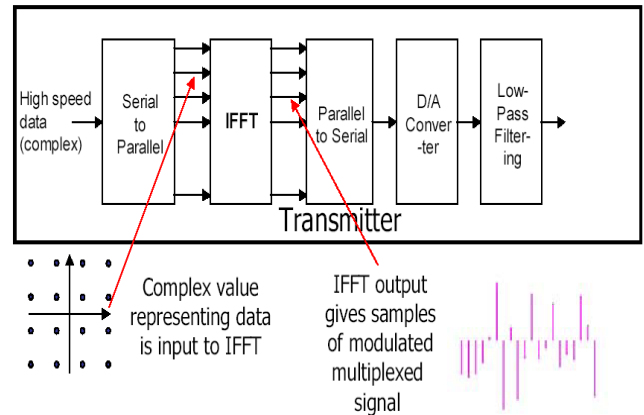


Fig. 1. OFDM transmitter

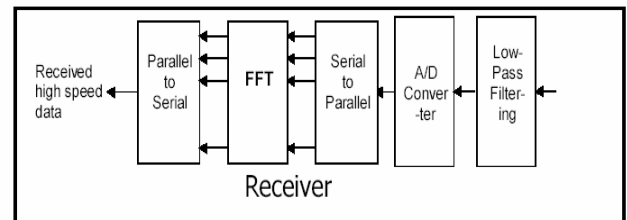


Fig 2 OFDM Receiver

After some additional processing, the time-domain signal that results from the IFFT is transmitted across the channel. At the receiver, an FFT block is used to process the received signal and bring it into the frequency-domain. Ideally, the FFT output will be the original symbols that were sent to the IFFT at the transmitter. When plotted in the complex plane, the FFT output samples will form a constellation, such as 16-QAM. However, there is no notion of a constellation for the time-domain signal. When plotted on the complex plane, the time-domain signal forms a scatter plot with no regular shape. Thus, any receiver processing that uses the concept of a constellation (such as symbol slicing) must occur in the frequency-domain.

3. SYNCHRONIZATION ISSUES

Synchronization has been one of the crucial research topics in orthogonal frequency division multiplexing (OFDM) system because of its sensitivity to the timing and frequency errors [2].

To guarantee the fast and accurate data transmission, the Inter Symbol Interference (ISI) and Inter Carrier Interference (ICI) caused in the transmission have to be eliminated as much as possible. In OFDM system, ISI can be avoided by inserting cyclic prefix with length greater than the channel impulse response, and the ICI can be eliminated by maintaining the orthogonality of carriers under the condition that the transmitter and the receiver have the exact same carrier frequency. But in the real world, frequency offsets will be arising from the frequency mismatch of the transmitter and the receiver oscillators and the existence of Doppler shift in the channel. In addition, due to the delay of signal when transmitting in the channel, the receiver in general starts sampling a new frame at the incorrect time instant. Therefore, it is important to estimate the frequency offset to minimize its impact, and to estimate the timing offset at the receiver to identify the start time of each frame and the FFT window position for each OFDM.

The OFDM synchronization can be divided into data-aided and non-data-aided categories[4]. The data-aided category uses a training sequence or pilot symbols for estimation. It has high accuracy and low calculation, but loses the bandwidth and reduces the data transmission speed. The non-data aided category often uses the cyclic prefix correlation. It doesn't waste bandwidth and reduce the transmission speed, but its estimation range is too small, not suitable for acquisition.

Three Synchronization Issues in the OFDM Systems

There are three major synchronization issues in the OFDM Systems:

- a. The symbol timing synchronization, which is to determine the correct symbol start position before the FFT demodulation at the receiver end.
- b. The carrier frequency synchronization (i.e., carrier frequency recovery technique), which is utilized to eliminate the carrier frequency offset caused by the mismatch from the local oscillators between the transmitter and the receiver, nonlinear characteristic of the wireless channel as well as the Doppler shift.
- c. The sampling clock synchronization, which is to mitigate the sampling clock errors due to the mismatch of the crystal oscillators.

All these synchronization errors will significantly degrade system performance

4. SYMBOL TIMING SYNCHRONIZATION

When signals are transmitted through severe channel conditions of multi-path fading, pulse noise disturbance and the Doppler Shift, it is important to solve symbol timing synchronization problem first during the design process of an OFDM receiver. The symbol timing error can not only disturb the amplitude as well as the phase of the received signal, but also introduce ISI.

In order to perform the FFT demodulation correctly, the symbol timing synchronization must be done to determine the starting point (i.e. FFT window) of the OFDM symbol. The cyclic prefix

(CP, or Guard Interval, GIB) can be removed afterwards. Accurate and steady symbol timing synchronization can be realized through the coarse symbol timing, the fine symbol timing as well as the symbol timing control structure combined together. The coarse symbol timing synchronization is first executed in time domain and then, the fine symbol timing in frequency domain is done to ensure a more accurate estimation. The symbol timing control structure is utilized to coordinate the operations of the coarse and the fine symbol timing.

5. SAMPLING CLOCK SYNCHRONIZATION

The sampling clock errors are mainly from the mismatch of the crystal oscillators between the transmitter and the receiver. Other factors such as multi-path fading, noise disturbance, Symbol timing estimation errors may also contribute to the sampling clock offset (SCO). The sampling clock errors will negatively influence the symbol timing synchronization

6. ADVANTAGES OF OFDM

OFDM has several advantages over single carrier modulation systems, some of these advantages are

1. **Multipath Delay Spread Tolerance:** OFDM is highly immune to Multipath delay spread that causes inter-symbol interference in wireless channels. Since the symbol duration is made larger (by converting a high data rate signal into 'N' low rate signals), the effect of delay spread is reduced by the same factor. Also by introducing the concepts of guard time and cyclic extension, the effects of inter-symbol interference (ISI) and inter-carrier interference (ICI) is removed completely.
2. **Immunity to Frequency selective fading Channels:** If the channel undergoes frequency selective fading, then complex equalization techniques are required at the receiver for single carrier modulation techniques. But in the case of OFDM the available bandwidth is split among many orthogonal narrowly spaced sub-carriers. Thus the available channel bandwidth is converted into many narrow flat-fading sub-channels. Hence it can be assumed that the sub-carriers experience flat fading only, though the channel gain/phase associated with the sub-carriers may vary. In the receiver, each sub-carrier just needs to be weighted according to the channel gain/phase encountered by it[5]. Even if some sub-carriers are completely lost due to fading, proper coding and interleaving at the transmitter can recover the user data.
3. **High Spectral Efficiency:** OFDM achieves high spectral efficiency by allowing the sub-carriers to overlap in the frequency domain. The sub-carriers are made orthogonal to each other therefore there is no Inter-Carrier Interference. If the number of sub-carriers is 'N', the total bandwidth required is $BW_{total} = (N+1)/T_s$. For large values of N, the total bandwidth required can be approximated as $BW_{total} = (N)/T_s$. On the other hand, the bandwidth required for single carrier transmission of the same data is $BW_{total} = (2N)/T_s$. Thus we achieve a spectral gain of nearly 100% in OFDM compared to the single carrier transmission case.
4. **Efficient Modulation and Demodulation:** Modulation and Demodulation of the sub-carriers is done using IFFT and FFT methods respectively, which are computationally efficient.
5. **Decrease complexity:** Key difference between single carrier modulation and ofdm is fft vs equalizer. Complexity of 64 point radix 4 fft in IEEE 802.11a 96 complex multiplication in four microsecond i.e. 96

million real multiplication per second. While 16 tap Gmsk equalizer at 24MHz means $2 \times 16 \times 24 = 768$ million real multiplications per second.

7. DISADVANTAGES OF OFDM

OFDM also have disadvantages over single carrier modulation systems, some of these disadvantages are:

1. The OFDM signal has a noise like amplitude with a very large dynamic range, therefore it requires RF power amplifiers with a high peak to average power ratio.
2. It is more sensitive to carrier frequency offset and drift than single carrier systems are due to leakage of the DFT.
3. High sensitivity to synchronization errors
4. Nonlinear effects generated by the power amplifier may introduce inter carrier interference and thus destroy the orthogonality
5. Larger sidelobes may result in sensitivity to frequency
6. It is highly vulnerable to synchronization errors.

8. CONCLUSION

In this paper, we focused on the major key synchronization issues in the OFDM systems. Typical algorithms such as the symbol timing, the carrier frequency and the sampling clock synchronization are discussed. Although research is going on to improve the performance of OFDM system.

REFERENCES

- [1] A. Czulwik, "Low overhead pilot-aided synchronization for single carrier modulation with frequency domain equalization," in *GLOBECOM*, vol. 4, 1998, pp. 2068–73.
- [2] H. Sari, G. Karam, and I. Jeanclaude, "Frequency-domain equalization of mobile radio and terrestrial broadcast channels," in *GLOBECOM*, 1994, pp. 1–5.
- [3] H. Sari, G. Karam, V. Paxal, and K. Maalej, "Trellis-coded constant envelope modulations with linear receivers," *IEEE Trans. Commun.*, vol. 44, pp. 1298–1307, Oct. 1996.
- [4] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Trans. Commun.*, vol. 43, pp. 191–193, Feb.–Apr. 1995.
- [5] J. V. de Beek, M. Sandell, and P. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 1800–1805, July 1997.
- [6] M. Morelli, A. Andrea, and U. Mengali, "Feedback frequency synchronization for OFDM applications," *IEEE Communications Letters*, vol. 5, no. 1, pp. 28–30, Jan. 2001.

-
- [7] U. Tureli, H. Lui, and M. O. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. on Communications*, vol. 48, no. 9, pp. 1459–1461, Sept. 2000.
- [8] U. Tureli and H. Liu, "Blind carrier synchronization and channel identification for OFDM communications," in *Proc. of IEEE ICASSP'98*, vol. 6, Seattle, WA, May 1998, pp. 3509–3512.
- [9] M. Luise, M. Marselli, and R. Reggiannini, "Low-complexity blind carrier frequency recovery for OFDM signals over frequency-selective radio channels," *IEEE Trans. on Communications*, vol. 50, no. 7, pp. 1182–1188, July 2002.
- [10] M. H. Hsieh and C. H. Wei, "A Low-complexity frame synchronization and frequency offset compensation scheme for OFDM systems over fading channels," *IEEE Trans. on Vehicular Technology*, vol. 48, no. 5, pp. 1596–1609, Sept. 1999.
- [11] M. Morelli, A. N. D'Andrea, and U. Mengali, "Frequency ambiguity resolution in OFDM system," *IEEE Commun. Letters*, vol. 4, no. 4, pp. 134–136, Apr. 2000.
- [12] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Communications Letters*, vol. 3, no. 3, pp. 75–77, Mar. 1999.
- [13] P. H. Moose, "A technique for orthogonal frequency-division multiplexing frequency offset correction," *IEEE Trans. on Communications*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.
- [14] B. Ai, J. H. Ge, Y. Wang, S. Y. Yang, P. Liu, and G. Liu, "Frequency offset estimation for OFDM in wireless communications," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 73–77, Mar. 2004.

A Review Paper on Zigbee

Amit Verma¹, Anvita Tripathi², Kapil Kumar³

^{1,2,3}Department of Electronics & Communication Engineering
Noida Institute of Engineering & Technology, Greater Noida, INDIA
¹amitverma0909@gmail.com, ²tripathi_ac@yahoo.com,
³kapil10305171mtech@gmail.com

Abstract: In this paper review of Zigbee technology, its transmission over MAC layer, various security issues & power consumption is discussed. Zigbee is a wireless network technology; this is based on IEEE 802.15.4 standard which provides low rate wireless personal area network (LR-WPAN). Zigbee includes four basic features i.e. low cost, low power consumption, low latency and short range communication for data exchange between devices. In this paper functionality of Zigbee at MAC and Physical layer with reference to architecture of IEEE 802.15.4 standard has been discussed in brief. Considering the security as an important feature in Zigbee products, basic security services at the MAC layer are also reviewed in this paper. The conclusion has been made with the comparison between Bluetooth and Zigbee technology.

Keywords: Zigbee; LR-WPAN; data communication; network co-ordinator; CSMA-CA;

1. INTRODUCTION

The name “Zigbee” has been given by an alliance of companies working together to enable Zigbee products based on an open global standard. Zigbee shares the 2.4 GHz band, which is an international ISM (industrial, scientific, medical) band particularly useful for device implementation [1]. This is already used by wireless LAN and Bluetooth. Unlike Bluetooth, Zigbee specifies a wireless system aimed at lower power consumption [2] and a data rate of up to 250 kbps[3]. Zigbee operates in two other unlicensed bands: 915MHz in North America and 868 MHz in Europe [10,11]. The former band has 10 channels with a maximum data rate of up to 40 kbps supported while the later has only 1channel that supports data rates up to 20 kbps [12, 29, 32].

Data security in Zigbee is achieved by using both symmetric and asymmetric key exchange protocols [4,5,6]. As a security issue, CSMA-CA channel mechanism is used for the transmission of data in beacon-enabled network [9,17,33,37]. Circular-Queue channel access mechanism shows better results with reduced collision rate, improved efficiency and improved performance in packet loss-rate [18]. IEEE standard 802.15.4 is utilised by Zigbee products to serve applications in wireless sensor networks (WSN)[7], monitoring, remote controls and sensors etc. where relatively low levels of data throughput are needed [20,21,25,41,49].

2. IEEE 802.15.4 STANDARD SPECIFICATION

The IEEE 802.15.4 standard is particularly designed for LR-WPAN based system. Some of the characteristics of LR-WPAN are –

- 1) Data rates of 250, 100, 40, 20 kbps.
- 2) Low power consumption.
- 3) Link quality indicator (LQI) [8,30].

Two major types of devices outlined in the IEEE 802.15.4 protocol are full function device (FFD) and reduced function device (RFD). FFD can operate in three modes – (a) PAN Co-ordinator (PC) (b) co-ordinator (c) node. A typical FFD device has the advantage to communicate with another FFD device or with a RFD device while RFD device can communicate only with a FFD device [9,52]. Thus FFD usually serves as a PAN co-ordinator (PC) and RFD fulfils the task of a nominal device such as a sensor only. In a Zigbee supported network, the PAN co-ordinator has the responsibility to start the network and choose key network parameters. IEEE 802.15.4 standard defines two layers-physical (PHY) layer, medium access control (MAC) layer.

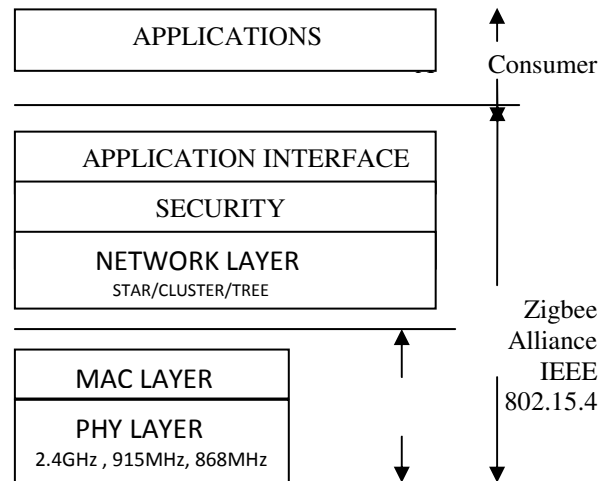


Fig. 1. Zigbee Protocol Stack [2]

Additional layers like network layer, an application layer and security services have been incorporated above the PHY layer and MAC layers that all together presents a complete Zigbee protocol stack [11, 31,37]. The layers are spaced

from each other by service access points (SAP). The network layer serves to provide security implementation and the communication of devices with the network co-ordinator.

3. NETWORK TOPOLOGY

The IEEE 802.15.4 standard offers 3 kinds of network topologies:

A. Star topology

In such a topology, one FFD acts as a PAN co-ordinator and all other devices are associated with it. Whenever a device wants to communicate with another device, it has to first send a message to the network co-ordinator.

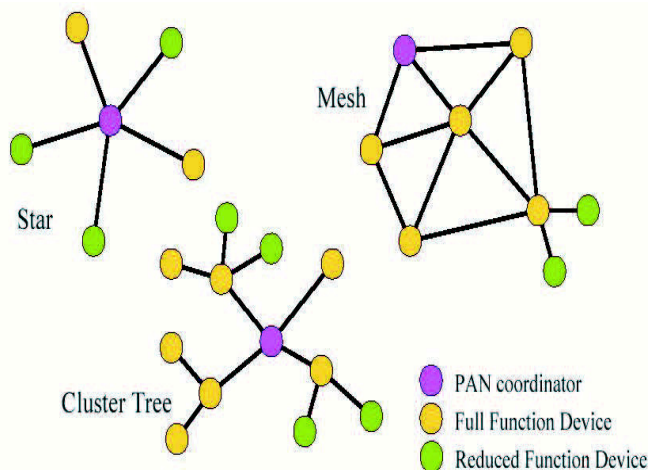


Fig. 2. Topology Models [15]

B. Mesh topology

In peer-to-peer topology, any device can communicate with any other device without the prior permission of the PAN co-ordinator. This topology uses the routing mechanism which would forward the message through a number of routing device to the destination source.

C. Cluster-tree topology

This is a special type of peer-to-peer network in which any device can operate as a FFD and provide synchronization to other devices or other co-ordinators. However, only one of these co-ordinators serves as a PAN co-ordinator. This topology shares the features of both the star topology and the mesh topology.

4. ANALYSIS OF THE MAC LAYER

The two basic services provided by the MAC sub layer are: the Mac data service and the Mac management service. The

various features of MAC sub layer include beacon transmission, synchronization, association and dissociation.

A. Data Transmission Model

Data is transmitted in Zigbee in 2 modes: beacon enabled and non-beacon enabled communication [13,34,35]. A beacon is defined as a packet containing information about the node and the network.

In beacon enabled network beacon frames are transmitted by the PAN at regular intervals of time. The beacons are used to synchronize the attached devices, to identify the network co-ordinator and to determine the structure of super frame [14,36,46] which has been described in detail in section IIIB. Due to this synchronization, the devices listen to the co-ordinator only during the beacon interval and go to sleep for the rest of time, thus conserving power. Non beacon-enabled network is generally preferred where the network co-ordinator does not operate on battery and hence there is no problem of power loss. For this kind of network the co-ordinator randomly sends beacon frames from time to time.

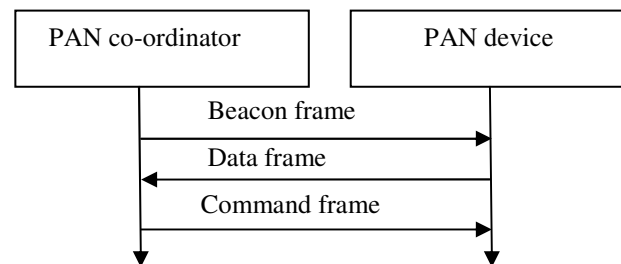


Fig. 3. Uplink Communication for beacon enabled network.[4]

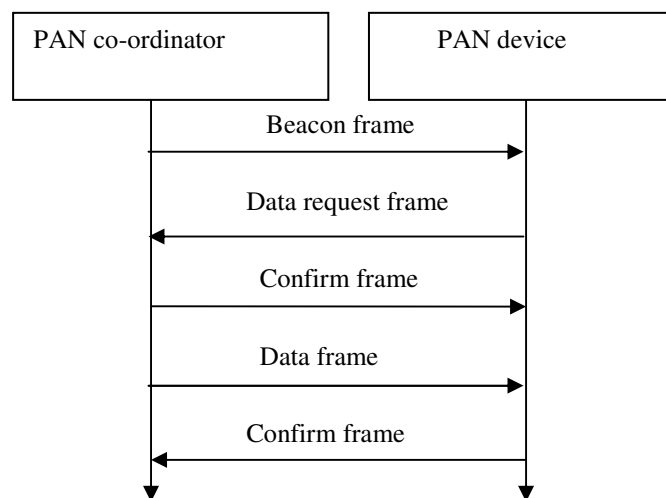


Fig. 4. Downlink Communication for beacon enabled network[4]

The attached devices have to wait for long periods of time and can only communicate with when the channel is idle for transmission or in a free state.

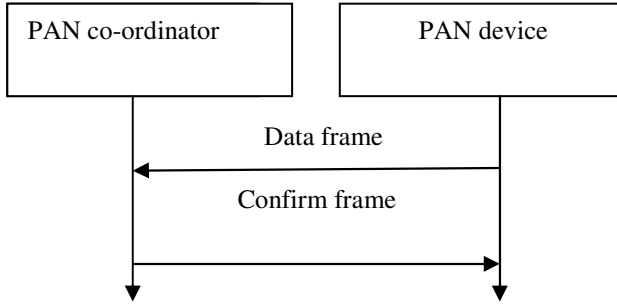


Fig. 5. Uplink Communication for non-beacon enabled network

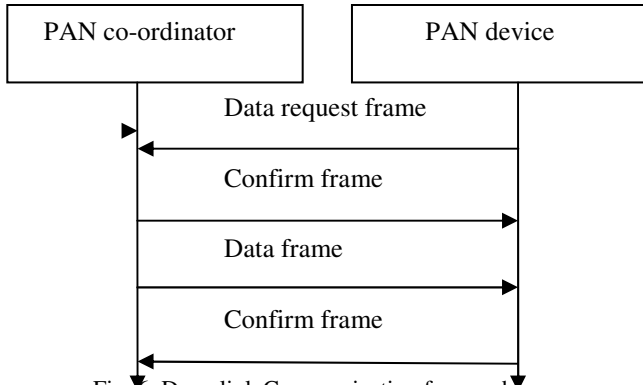


Fig. 6. Downlink Communication for non-beacon enabled network [4]

B. Super frame Structure

When the network is operating in beacon-enabled mode, the channel-time is divided into super frames. A superframe structure comes into action only when data is transmitted using the beacon-enabled mode. The super frame structure is bounded by two beacon frames. It is not mandatory to use the superframe structure in the IEEE 802.15.4 standard. Network co-ordinator decides the format of a superframe structure [15]. No beacons are transmitted when the co-ordinator does not use the super frame format. The communication in a beacon-enabled network takes place during the active portion of the superframe only. There can be 3 types of periods in a superframe structure:

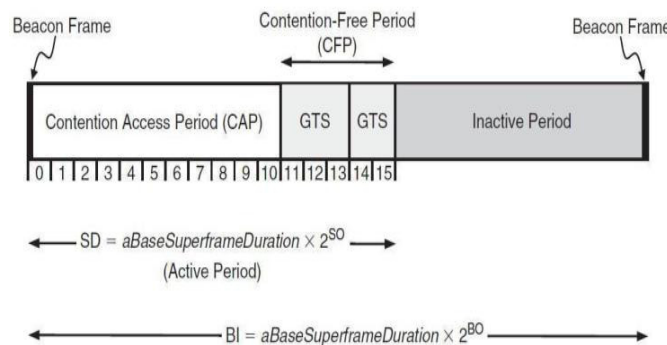


Fig. 7. Superframe Structure [15]

- i. Contention Access Period (CAP)
- ii. Contention Free Period (CFP)
- iii. Inactive Period

The combination of CAP and CFP is known as the active period. The super frame structure set up by the network co-ordinator is used to control the channel access. The important parameters defined in the super frame structure are: macBeaconOrder (BO) and macSuperFrameOrder (SO) [18]. The interval that is used to transmit the beacons is called the beacon-interval. Beacon interval, BI is related to macBeaconOrder as follows: $BI = aBaseSuperFrameDuration \times 2^{BO}$, $0 \leq BO \leq 14$.

The length of the active portion of the super frame is known as the Super frame Duration. The relationship between the macSuperFrameOrder SO is given as: $0 \leq SO \leq 14$ [15]. No beacons are transmitted by the co-ordinator when the values of macBeaconOrder and macSuperFrameOrder are set to 15 and the GTS shall not be permitted in this state [16].

C. Beacon transmission

Full function device (FFD) may either operate in the beacon-enabled mode or a co-ordinator or as a device on a pre-established co-ordinator or in a non beacon-enabled network. Network co-ordinator starts transmitting beacons when it wants to communicate with the attached devices on a network but an FFD that is not a co-ordinator can only transmit beacons after successful association with network co-ordinator. Beacons are transmitted periodically and provide 16 equal-width time slots [15] between beacons for contention free channel access in each time slot. The macBeaconOrder and macSuperFrameOrder also need to be taken care while dealing with a beacon-enabled network. mac BeaconTxTime stores the time of transmission of the most recent beacon on the network.

D. Channel access

When a device on the network learns that there is a pending message, it sends a request to the co-ordinator. The data is sent by the co-ordinator upon accepting the request. Since both packet transmission and clear channel access (CCA) checks must be synchronized with the slot-boundaries of the back-off periods, the channel access mechanism is known as carrier-sense medium access-collision (CSMA-CA) [17]. When the beacons are transmitted by the network co-ordinator in a super frame structure, then slotted (CSMA-CA) algorithm is used. Each device on the network has 3 variables: NB, CW and BE. NB is the number of times the CA algorithm was required to back-off while attempting the current transmission. NB holds 0 as the initial value. CW is the contention window that defines the number of back-off periods cleared before transmission. CW is initiated to 2 before a new start. BE is the back-off exponent, defined as

how many back-off periods a device shall wait before assessing the channel [16].

E. MAC frame structure

Every frame structure consists of:

- (i) MAC frame header (MHR), which include frame control, sequence number and address information.
- (ii) MAC frame payload (MAC payload) of variable length containing specific frame information.
- (iii) MAC frame tail (MFR) that contains FCS.

Table 1. Frame Structure of Mac Layer

2 bytes	Frame control information
1 bytes	Frame serial number
0/2 bytes	Destination device PAN identifier
0/2/8 bytes	Destination address
0/2 bytes	Source device PAN identifier
0/2/8 bytes	Source address
Variable length	Frame payload
2 bytes	Frame check

There are 4 frame types in MAC layer-

Beacon frame: The sub- field of the frame type is set to (000). This frame is used to add a new level of functionality in the network.

Command frame: This frame serves to provide network connection, disconnection and channel access. This frame controls the working of client nodes. The structure of the command frame is the most complex.

Data frame: It ensures that all the data packets are tracked. This frame is relatively simple, where the sub-domain of the frame type is set to data frame type (001). The frame payload contains the information coming from upper layers.

Response frame: This is the simplest frame in the mac frame structure. The sub-domain of the frame type is assigned the response frame type (010) Provides the feedback from receiver to sender that the packets are received without error.

F. Synchronization

The synchronization in Zigbee network can be achieved in 2 ways: beacon-frame or pulling data mechanism.

1) Synchronization mechanism

In a super frame structure beacons are periodically transmitted by the network co-ordinator. The devices on the network track the beacon-frames and verify whether the source in address of the beacon matches with the co-ordinator address, if yes, then beacon is accepted otherwise rejected. The data from the accepted beacon is used to communicate in the next super frame and this is how synchronization is attained using beacon mechanism.

2) Pulling data mechanism

This mode of synchronization is used in a network that does not support super frame structure. Whenever the node learns that there is a pending message, it sends a request command frame to pull up data from the network co-ordinator [19]. The device has to wait for a long period of time till the request is accepted and the data is send by the co-ordinator.

5. ZIGBEE VS. BLUETOOTH

From the inception of Zigbee there has been a tussle between the experts of both wireless technologies. The following table easily differentiates the two:

Table II. Comparison between Bluetooth & Zigbee

Market Name	Bluetooth	Zigbee
Standard	802.15.1	802.15.4
Application	Cable	Monitoring &
Focus	Replacement	Control
Battery Life	1-7	100-1000+
(Days)	7	255/65,000
Network Size	1 Mbps	250Kbps
Data Rate	1-10+	1-100+
Transmission	Cost,	Reliability,
Range(Metres)	Convenience	Power, Cost
Success Metrics		

6. POWER CONSUMPTION

Power consumption is a key aspect of Zigbee technology to allow long lifetime for the Zigbee based products. Power optimization for the Zigbee networks is done for the slave nodes [7]. The strategy for the energy conservation in the communication module mainly comprises of the communication modulation, increasing the sleeping time and the wireless communication method of multihopping short range [53].

7. SECURITY SERVICES

Data security is an important issue in networking systems [28, 38, 39]. IEEE standard 802.15.4 does not provide any

procedure for key management or device authentication [40,42]. Hence these services must be implemented by another layer of network protocols running on top of 802.15.4 itself [4,5,47,53].

The combined work performed by the Zigbee Alliance and the 802.15.4 group meets the security requirements in the industrial environments. These devices are programmed to perform specific tasks or provide specific information accurately and reliably [47]. Security services are provided at different layers using both asymmetric and symmetric key-exchange protocols.

A. Asymmetric key-exchange protocol:

This protocol is designed specifically for devices with high bandwidth links, rich in power and computation sources. This protocol mainly rely on public-key cryptography[4]. The growing wireless sensor networks may have a large number of nodes, hence the security and management will present a crucial requirement. Therefore, it is essential to implement the public-key cryptography which identifies every single node on the network and enables security of data at the transmission and the receiving end. In public-key cryptography, one key binds the device on the network while the other key is used to verify that identity [53]. Hence, in this manner the identification of the device on the network is performed rapidly and in a strong manner.

B. Symmetric key- exchange protocol:

The networks that require low power operations based on simple, resource-constrained nodes favour the use of symmetric key-exchange protocols. One such protocol is the symmetric key key establishment (SKKE) protocol [4]. In such a protocol, one Zigbee device shares a secret key with another Zigbee device. This establishment supports two quantities namely an initiator and a responder device. Zigbee uses the Advanced Encryption Standard (AES) for the symmetric encryption because it is faster and also inexpensive.

8. FUTURE ASPECTS

Ever since the date of Zigbee's arrival, it has been able to gain a lot of ground in the market and keen to play a huge role in future wireless technology. Wireless Sensor Networks (WSNs) have emerged as the leading technologies with a future to combine automated sensing embedded computing and wireless networking into tiny embedded devices [24]. The upcoming Zigbee products are expandable in applications to wireless sensors for the security and control of home and office Buildings, remote control for the industry and consumer electronics.

REFERENCES

- [1] Chris Evans-Pughe, "Is the Zigbee wireless standard, promoted by an alliance of 25 firms, a big threat to Bluetooth?" IEE review, volume 49, pp.28-31, March 2003.
- [2] Mikhail Galeev, "Making Friends with Zigbee: Introduction to 802.15.4 Protocol", Emerson Network Power, Embedded System Conference, Boston, June 2007.
- [3] Ian Poole, "What Exactly Is...Zigbee?", IET, Communication Engineer, volume 4, pp.44-45, August-September 2004.
- [4] Jelena Mistic, "Zigbee: a long way to go", Elsevier Ltd, Info Security, volume 4, issue 3, pp.32-35, 2012.
- [5] Mitch Blazer, "Industrial-Strength Security for Zigbee: The case for Public-key Cryptography", Embedded Computing Design, May 2005.
- [6] Biming Tian, Sung Han, Liu Liu, Saghar Shadem, Sazia Parvin, "Towards enhanced key management in multi-phase Zigbee network architecture", Elsevier Ltd, Computer Communications, volume 35, issue 5, pp. 579-588, March 2012.
- [7] Shizhong Chen, Jinmei Yao, Yuhou wu, "Analysis of the Power Consumption for Wireless Sensor Network Node Based on Zigbee", Elsevier Ltd, Procedia Engineering, International Workshop on Information and Electronics Engineering (IWIEE), volume 29, pp. 1994-1998, 2012.
- [8] Abhinav Singh, Pankaj Kumar Patel, P. C. Jain, "Performance Analysis of Beacon Enabled IEEE 802.15.4 using GTS in Zigbee WPAN using Zigbee", IJERA, vol 2, issue 4, pp.461-465, July-August 2012.
- [9] Zigbee Specifications, www.Specifications.net/zigbee/w_UK.php.
- [10] Cliff Bowman, "Applying Embedded Wireless Networks", Embedded Systems Conference, Massachusetts, class#349, June 2004.
- [11] Christopher Leidigh, Jim Higgins, "Inside Look at Zigbee Wireless Networks", ESC-252, Embedded Systems Valley, April 2007.
- [12] Nick Hunn, "Choosing a Short Range Wireless Technology", Ezurio ltd, white paper, 2005.
- [13] Min Zhou, Hang-long Nie, "Analysis and Design of Zigbee MAC layer protocol", IEEE, International Conference on Future Information Technology and Management Engineering, volume 2, pp. 211-215, oct 2010.
- [14] Abhinav Singh, Pankaj Kumar Patel, P. C. Jain, "Performance Analysis of Beacon Enabled IEEE 802.15.4 using GTS in Zigbee", International Journal of Engineering Research and Applications (IJERA), ISSN:22248-9622, volume 2, issue 4, pp. 461-465, July/August 2012.
- [15] Abhinav Singh, Pankaj Kumar Patel, P. C. Jain, "Analysis of Throughput using GTS and Multichannel in IEEE 802.15.4", International Journal of Advanced Technology & Engineering Research (IJATER), ISSN: 2250-3536, vol.2, Issue 2, pp.44-48 March 2012.
- [16] Abhinav Singh, Pankaj Kumar Patel, P. C. Jain, "Analysis of Throughput GTS and SDS in IEEE 802.15.4", International Journal of Electronics and Computer Science Engineering, ISSN: 2277-1956, vol.1, number 2, pp.597-602, 2012.
- [17] Jelena Mistic, Shairmina Shafi, Vojislav b. Mistic, "Performance Limitations of the MAC layer in 802.15.4 low rate WPAN",

- ACM, Journal of Computer Communications, volume 29, issue 13-14, pp. 2534-2541, august 2006.
- [18] Shi Long Long, Qiu Chunling, Gau Peng, Jia Zhengsen, "The Research and Simulation of CSMA-CA Mechanism of Zigbee Protocol", Elsevier Ltd, Procedia Engineering, International Workshop on Information and Electronics Engineering (IWIEE), volume 29, pp. 3466-3471, 2012.
- [19] Chia-Mingwu, Ruay-Shiung Chang, "An Innovative Scheme for Increasing Connectivity in Zigbee Networks", IEEE, International Conference on Parallel Processing, pp. 99-104, sept 2011.
- [20] Drew Gislason, "Zigbee Wireless Networking", Elsevier Inc, 1st edition, ISBN: 978-0750685979, 2007.
- [21] Tim Gillman, Drew Gislason, "Application Development for Wireless Networking", Embedded System Conference, ETP-449, March 2005.
- [22] www.rfide.com
- [23] www.stg.com
- [24] M. Aykut Yigitel, Ozlem Durmaz Incel, Cem Ersoy, "QoS-Aware Mac Protocols for wireless Sensor Networks: a survey", Elsevier Ltd, Computer Networks, volume 55, issue 2, pp. 1982-2004, 2011,
- [25] pages.cs.wisc.edu/~suman/courses/838/papers/zigbee.pdf
- [26] B. E. Bilgin, V.C Gungor, "Performance Evaluations of Zigbee in Different Smart Grid Environments", Elsevier Ltd, Computer Networks, volume 56, issue 8, pp. 2196-2205, 2012. [27] N. Vljajic, D. Stevananovic, G. Spangiann Opoulus, "Strategies for Improving Performance Constrained Mobile sink(s)", Elsevier Ltd, Computer networks, volume 34, issue 6, pp. 743-757, 2011.
- [27] Maoheng Sun, "Study and Applications of security based on Zigbee standard", IEEE, International Conference on Multimedia Information Networking and Security (MINES), 2011, pp. 508-511, 2011.
- [28] Paolo Baronti, Prashant Pillai, Vince W.C. Chook, Stefanoc Chessa, Alberto Gotta, Y. Fun Hu, "Wireless sensor networks: A survey on the state of the art and 802.15.4 and Zigbee standards", Elsevier Ltd, Computer communications, volume 30, issue 7, pp. 1655-1695, May 2007.
- [29] Yan Xin, Huahua Yao, Yingying Jiang, Shu Yan, Jun Sun, "Analysis and Design of ZigBee Network Layer Protocol under Cellular Network Environment", IEEE, International Conference on Computer Science and Electronics Engineering, volume 2, pp. 59-62, 2012.
- [30] Shahin Farahani, "Zigbee Wireless Networks and Transreceivers", Elsevier Inc, 2008, ISBN: 978-0-7506-8393-7.
- [31] Chihhsiong Shih, Bwo-cheng Liang, "A Model Driven Software Framework for Zigbee Based Energy Saving Systems", IEEE, Third International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 487-492, 2012.
- [32] Shahin Farahani, "Hello Zigbee", Zigbee Wireless Networking, Elsevier Inc, 1st edition, pp. 1-44, ISBN: 978-0750685979, 2008.
- [33] Jelena Mistic, Shairmina Shafi, Vojislav B. Mistic, "The impact of MAC parameters on the performance of 802.15.4 PAN", Elsevier Inc, Ad Hoc Networks, Volume 3, Issue 5, pp. 509-528, September 2005.
- [34] Kurtis Kredo, Prasant Mohapatra, "Medium access control in wireless networks", Elsevier Inc, Computer Networks, volume 51, issue 4, pp. 961-994, March 2007.
- [35] Helena Fernandez-Lopez, Jose A. Afonso, J.H. Correia, Ricardo Simoes, "Towards the Design of Efficient NonBeacon-Enabled Zigbee Networks", Elsevier Inc, Computer Networks, volume 56, issue 11, pp. 2714-2725, July 2012.
- [36] Hua Qin, Wensheng Zhang, "Zigbee-Assisted Power Saving Management for Mobile Devices", Department of Computer Science, Iowa State University, USA
www.cs.iastate.edu/~qinhua/papers/MASS12.pdf
- [37] Vadym Samosuyev, "Bluetooth Low Energy Compared to Zigbee and Bluetooth Classic", Bachelor's Thesis, Information Technology, May 24 2012 https://publications.theseus.fi/bitstream/handle/10024/15812/FinalThesis_Samosuyev.pdf
- [38] Jiho Kim, "Power Efficient Architecture of Zigbee Security Processing", IEEE, International Symposium on Parallel and Distributed Processing with Applications, pp. 773-778, 2008.
- [39] Li Chunqing, Zhang Jiancheng, "Research of Zigbee's Data Security and Protection", IEEE, International Forum on Computer Science-Technology and Applications, volume 1, pp. 298-302, 2009.
- [40] Gianluca Dini, Marco Tiloca, "Considerations on Security in ZigBee Networks," *sutc*, pp.58-65, 2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, 2010
- [41] Bin Yang, "Study on Security Wireless Sensor Network Based on Zigbee Standard", IEEE, International Conference on Computational Intelligence and Security, volume 2, pp. 426-430, 2009.
- [42] Meng Qianqian, Bao Kejin, "Security Analysis for Wireless Networks Based on ZigBee," *ifita*, vol. 1, pp.158-160, 2009 International Forum on Information Technology and Applications, 2009
- [43] Min Zhou, Zhang-long Nie, "Analysis and Design of Zigbee MAC layers protocol", International Conference on Future Information Technology and Management Engineering (FITME), volume 1, pp. 211-215, 2010.
- [44] Mohana, P.Radha, "Realization of MAC layer functions of Zigbee protocol Stack in FPGA", International conference on Control, Automation, Communication and energy conservation (INCACEC), pp. 1-5, 2009.
- [45] Sarvakar K, Patel P.S, "An efficient hybrid MAC layer protocol utilised for wireless sensor networks", Fourth International Conference on Wireless Communication and Sensor Networks, pp. 22-26, 2008
- [46] Zhang X, Shin K.G, "Cooperative Carrier Signalling: Harmonizing Coexisting WPAN and WLAN devices", IEEE/ACM Transactions on Networking, volume PP, issue 99, pp. 1, 2012.
- [47] Khan S A, Khan F.A, "Performance Analysis of a Zigbee Beacon Enabled Cluster Tree Network", IEEE, Third International Conference on Electrical engineering, pp. 1-6, 2009.
- [48] Meng Qianqian, Bao Kejin, "Security Analysis for Wireless Networks Based on ZigBee," *ifita*, vol. 1, pp.158-160, 2009 International Forum on Information Technology and Applications, 2009

- [49] Adams, J.T., "An Introduction to IEEE STD 802.15.4", IEEE Conference on Aerospace, 2006
- [50] Chih-Kuang Lin, Kokkinos T, Mullany F, "Extended-range Wireless Sensor Networks with enhanced IEEE 802.15.4 CSMA/CA ", IEEE Sensors, ICSENS, pp. 1994-1997, 2011.
- [51] Hongwei Li, Zhongning Jia, Xiaofeng Xue, "Application and Analysis of Zigbee Security Services Specification", IEEE, Second International Conference on Networks Security Wireless Communication and Trusted Computing (NSWCTC), volume 2, pp. 494-497, 2010.
- [52] Amr Amin Hafez, Mohammed Dessovsky, Fikri Ragai, "Design of low power Zigbee receiver front-end for wireless sensors", Elsevier ltd., 1561-1568, 2009.

New Emerging Era of Communication: Broadband Technology - A Review

Mitesh Sharma

*Assistant Professor
Department of Computer Science
Jodhpur Institute of Engineering & Technology, Jodhpur
mail2miteshsharma@gmail.com*

Abstract: As broadband becomes more widely diffused in world. There is great potential to increase the number of people who are connected via broadband. Broadband wireless access networks are considered to be enterprise-level networks providing more capacity and coverage. Wireless networking has offered an alternative solution to the problem of information access in remote inaccessible areas where wired networks are not cost-effective. They have changed the way people communicate and share information by eliminating worrisome factors of distance and location. This paper provides an overview of broadband technologies with bandwidth management and use via mobile also.

Keywords: Broadband, bandwidth management, benefits etc.

1. INTRODUCTION

Broadband communication is becoming a foundational element of the entire economy, supporting entire industries, transforming not only how people work, but how they lead their lives. As wireless technology represents an increasing portion of the global communications infrastructure, it is important to understand overall broadband trends. Sometimes wireless and wireline technologies compete with each other, but in most instances, they are complementary. For the most part, backhaul transport and core infrastructure for wireless networks are based on wireline approaches, whether optical or copper. This applies as readily to Wi-Fi networks as it does to cellular networks.

The Industrial Revolution during the past two centuries produced the most development in the history of mankind. But that period of unparalleled growth will be overshadowed by the current technological revolution, namely, the Information and Communication Technology (ICT) revolution. This revolution will not only benefit individual citizens but will have a tremendous impact on national economies and the global economy as a whole. As a result of ever-increasing global connectivity, the amount of information that can be transmitted electronically has grown exponentially, resulting in unprecedented ease of communication in most of the countries.

To realize the true success of the ICT revolution, broadband connectivity is needed, as it is not only information that is shared but also voice, images, video, etc. There is no agreed definition of broadband, but it is usually recognized by its higher transmission speed and “always-on” connectivity. Broadband is at the heart of the convergence of telecommunication, information technology, and broadcasting. Therefore, there is a great need for modern high-tech communication infrastructure since the focus of applications is on interactivity rather than just information sharing.

There are multiple factors contributing to explosive growth in data consumption, but first and foremost is the combination of powerful mobile computing platforms and fast mobile broadband networks. Despite the number of vendors and platform types available on the device side, the industry is converging on what might be considered a “standard” platform for smart-phones and also one for tablets. Even if implemented differently, these platforms have the capabilities shown in Figure 1.

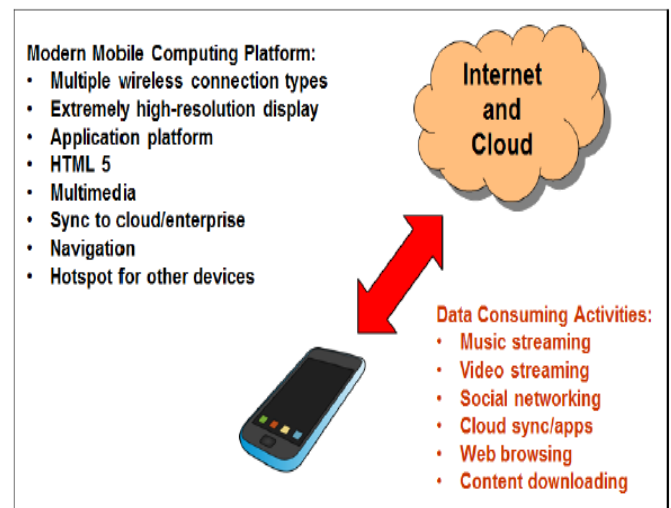


Fig. 1. Modern Mobile Computing Platform and Data Consumption

The rich capabilities of these mobile platforms enable them to consume ever larger amounts of data through activities such as music and video streaming, social networking, cloud-based synchronization and applications, Web browsing, and content downloading.

Table 1: Data Consumed by Different Streaming Applications

Application	Throughput (Mbps)	MByte/hour	Hrs./day	GB/month
Audio or Music	0.1	58	0.5	0.9
			1.0	1.7
			2.0	3.5
			4.0	6.9
Small Screen Video (e.g., Feature Phone)	0.2	90	0.5	1.4
			1.0	2.7
			2.0	5.4
			4.0	10.8
Medium Screen Video (e.g., Smartphone Full-Screen Video)	1.0	450	0.5	6.8
			1.0	13.5
			2.0	27.0
			4.0	54.0
Larger Screen Video (e.g., Netflix Lower Def. on Tablet or Laptop)	2.0	900	0.5	13.5
			1.0	27.0
			2.0	54.0
			4.0	108.0
Larger Screen Video (e.g., Netflix Higher Def. on Laptop)	4.0	1800	0.5	27.0
			1.0	54.0
			2.0	108.0
			4.0	216.0

With declining voice revenue, but increasing data revenue, cellular operators face a tremendous opportunity in continuing to develop their mobile broadband businesses. Successful execution, however, means more than just providing high speed networks. It means addressing demand that is growing at an extremely rapid rate. It also means nurturing an application ecosystem, delivering complementary services, providing a compelling customer experience, and supplying attractive devices. These are all areas in which the industry has done well.

2. WIRELESS VERSUS WIRELINE

Wireless technology is playing a profound role in networking and communications, even though wireline technology such as fiber has inherent capacity advantages. The overwhelming global success of mobile telephony and now the growing adoption of mobile data conclusively demonstrate the desire for mobile-oriented communications. Mobile broadband combines compelling high-speed data services with mobility. Thus, the opportunities are limitless when considering the many diverse markets mobile

broadband can successfully address. Developed countries continue to show tremendous uptake of mobile broadband services. Additionally, in developing countries, there is no doubt that 3G and 4G technology will cater to both enterprises and their high-end mobile workers and consumers for whom mobile broadband can be a cost-effective option competing with digital subscriber line (DSL) for home use.

Relative to wireless networks, wireline networks have always had greater capacity and historically have delivered faster throughput rates. Even if mobile users are not streaming full-length movies in high definition, video is finding its way into an increasing number of applications including education, social networking, video conferencing, business collaboration, field service, and telemedicine. Over time, wireless networks will gain substantial additional capacity through all the methods discussed in the next section, but they will never catch up to wireline. One can understand this from a relatively simplistic physics analysis:

- Wireline access to the premises or to nearby nodes uses fiber-optic cable.
- Capacity is based on available bandwidth of electromagnetic radiation. The infra-red frequencies used in fiber-optic communications have far greater bandwidth than radio.
- The result is that just one fiber-optic strand has greater bandwidth than the entire usable radio spectrum.

3. BROADBAND TECHNOLOGY

Broadband Technologies are very useful in all respect of life utilities. Following are the recent technologies:

- Wi-Fi:** Wi-Fi is the first high-speed wireless technology to enjoy broad deployment, most notably in hotspots around the world – including homes and offices, and increasingly cafes, hotels, and airports. In specification Wi-Fi hotspots became popular almost immediately and have been applauded by road warriors for their ability to improve productivity. Wi-Fi is limited, however, by its range: high-speed connectivity is possible only as long as a user remains within range of the wireless access point, which is optimum within 300 feet. Wi-Fi was one of the earliest high-speed wireless data technologies and now benefits from a broad availability of supporting products and technologies. Intel Centrino mobile technology optimizes performance in mobile data platforms, helping users get the most from the expanding Wi-Fi infrastructure. Some of the newest platforms even support multiple Wi-Fi standards (e.g. 802.11a, b and/or g) for compatibility among several wireless networks.

- b) **WiMAX:** WiMAX is an emerging technology that will deliver last mile broadband connectivity in a larger geographic area than Wi-Fi, enabling T1 type service to business customers and cable/DSL-equivalent access to residential users. Providing canopies of coverage anywhere from one to six miles wide (depending on multiple variables), WiMAX will enable greater mobility for high-speed data applications. With such range and high throughput, WiMAX is capable of delivering backhaul for carrier infrastructure, enterprise campuses and Wi-Fi hotspots. WiMAX will be deployed in three phases. Phase one will see WiMAX technology using the IEEE 802.16d specification deployed via outdoor antennas that target known subscribers in a fixed location. Phase two will roll out indoor antennas, broadening the appeal of WiMAX technology to carriers seeking simplified installation at user sites. Phase three will launch the IEEE 802.16e specification, in which WiMAX-Certified* hardware will be available in portable solutions for users who want to roam within a service area, enabling more persistent connectivity akin to Wi-Fi capabilities today.
- c) **3G:** 3G is an ITU specification for high-speed wireless communications. This worldwide wireless connection is compatible with GSM, TDMA, and CDMA. Next-generation 3G cellular services will provide a long-range wireless access canopy for voice and data. Carriers worldwide are now in the process of deploying 3G network infrastructure across urban, suburban and highly trafficked rural areas. Next-generation 3G cellular services will create broad-range coverage for data access across wide geographic areas, providing the greatest mobility for voice communications and Internet connectivity. 3G services will enable highly mobile users with laptops and other wireless data devices to bridge the gap between higher bandwidth WiMAX hot zones and Wi-Fi hotspots. New devices optimized for 3G communications are beginning to reach the marketplace. Such devices include cell phones that can also provide interactive video conferencing, as well as PDAs that can provide full-playback DVD services. 3G technologies are designed to provide the greatest mobility and are intended for devices whose primary function is voice services with additional data applications as a complement to those services.
- d) **UWB:** Ultra-Wideband (UWB) is a future wireless personal area network (WPAN) technology capable of high throughput (up to 400Mbps) at very short range (less than 30 feet). UWB will likely be utilized to enable wireless USB access for connecting computer peripherals to a PC and multiple components in the consumer electronics stack – e.g. home theater

equipment. UWB has the throughput capability to simultaneously distribute multiple high definition video streams. Intel engineers are working with a variety of industry leaders to develop a standard UWB radio platform. Made up of two core layers – the UWB radio layer and the convergence layer – the UWB platform will serve as the underlying transport mechanism for different applications that would operate on top of the single radio, such as wireless universal serial bus (USB), IEEE 1394, the next generation of Bluetooth and Universal Plug and Play.

4. THE NEED FOR BROADBAND WIRELESS IN DISASTER AND EMERGENCY RESPONSE

Broadband wireless connectivity can provide significant capabilities at a disaster or emergency site to increase the safety of the responders and to increase the effectiveness of the response. A number of related factors are increasing the need for high data rate network connectivity for disaster and emergency response. Brief summary of these factors are given below:

- The disaster and emergency management community has, over time, developed effective processes to respond to crises. An Incident Command System (ICS), customized to the scale and nature of the event, is used by federal, state, and local agencies responding to a crisis.
- Existing communications support is focused on voice. Land mobile radio (LMR) systems can provide local and regional connectivity, while cellular telephone systems can provide national connectivity. However, data connectivity is becoming increasingly important as part of the IT infrastructure required for the ICS.
- Communications is required locally within the disaster site, within the local region, and nationally (or even globally in some cases). Thus, there is a need for at least three tiers of communications infrastructure:
 - i. local connectivity, e.g., using wired and wireless local area network (LAN) technology;
 - ii. backbone or backhaul connectivity, which is the focus of our work;
 - iii. wide area network (WAN) connectivity in the form of the global Internet or a private network. The backbone network may connect directly to the WAN or a fourth component – realized using satellite communications or terrestrial wireless – may be needed to connect the backbone to the WAN.
- Responders cannot rely solely on the public communications infrastructure. It may be destroyed or largely destroyed by the disaster, it may be non-existent

as is the case in many rural areas, or it may be saturated by public safety users, the press, and others.

- There is an increasing need for interoperability to support multi-agency response. Large scale disasters require response by agencies in adjoining jurisdictions. Many incidents also require the involvement of multiple federal, state, and local agencies with different charters. The lack of interoperability is most evident today with LMR systems where, for example, fire fighters from one city cannot communicate directly with fire fighters from an adjoining city who are trying to assist.
- The communications infrastructure must be flexible to satisfy a variety of situations. Different locations and different types of emergencies and disasters may require use of different applications, connectivity for different types of end-user equipment, support for different types of users, operation in different environments, etc. Clearly, all equipment used for emergency and disaster response must be rugged to survive transport and harsh conditions and easy to use by responders who need technology to be “transparent” so that they may focus on life-critical tasks. It is also important to note that responders are focused on their immediate and critical mission. Technology that shows clear, immediate, and significant benefits will likely be adopted. Technology that is confusing, ineffective, or requires significant training will likely be ignored.

5. BANDWIDTH MANAGEMENT

Given huge growth in usage, mobile operators are either employing or considering multiple approaches to manage bandwidth:

- **More spectrum.** Spectrum correlates directly to capacity, and more spectrum is becoming available globally for mobile broadband. In the U.S., the FCC National Broadband Plan seeks to make an additional 500 MHz of spectrum available by 2020.
- **Unpaired spectrum.** Technologies such as HSPA+ and LTE allow the use of different amounts of spectrum between downlink and uplink. Additional unpaired downlink spectrum can be combined with paired spectrum to increase capacity and user throughputs.
- **Increased spectral efficiency.** Newer technologies are spectrally more efficient, meaning greater aggregate throughput in the same amount of spectrum. Wireless technologies such as LTE, however, are reaching the theoretical limits of spectral efficiency and future gains will be quite modest, allowing for a possible doubling of LTE efficiency over currently deployed versions.
- **Combining uplink gains with downlink carrier aggregation.** Operators can increase network capacity

by applying new receive technologies at the base station (e.g., large scale antenna systems) that do not necessarily require standards. This can be combined with added capacity on the downlink from carrier aggregation. This type of deployment flexibility suggests that regulators should consider licensing just downlink spectrum in some cases, since that is where it is generally most needed.

- **Wi-Fi.** Wi-Fi networks offer another means of offloading heavy traffic, especially as the number of Wi-Fi hotspots increases and connections become more seamless. Wi-Fi adds capacity since it offloads onto unlicensed spectrum. Moreover, since Wi-Fi signals cover only small areas, Wi-Fi achieves both extremely high frequency re-use, as well as high bandwidth per square meter across the coverage area.
- **Off-peak hours.** Operators can offer user incentives or perhaps fewer restrictions on large data transfers that occur at off-peak hours such as overnight.
- **Quality of service (QoS).** By prioritizing traffic, certain traffic such as non-time-critical downloads can execute with lower priority, thus not affecting other active users.
- **Innovative data plans.** Creative new data plans influencing consumption behavior including tiered pricing make usage affordable for most users, but discourage excessive or abusive use.
- **Explore new methods for the future.** Recently there has been a considerable amount of discussion about spectrum sharing. Although a promising approach for better spectrum utilization in the long term, spectrum sharing will require new technologies, as well as spectrum coordination, items that could take ten years or more to develop and commercialize.

It will take a creative blend of all of the above to make the mobile broadband market successful and to enable it to exist as a complementary solution to wired broadband.

6. BENEFITS

To ensure that the goal for access to broadband is met, the Federal Communications Commission (FCC) is reaching out to businesses, consumers and municipalities alike so that no community is left behind. Broadband deployment is so crucial because of the many ways it touches people’s lives. Broadband services provide users high-speed access to data, video, audio and voice services all over one connection and bringing tremendous benefits and achieving important goals.

- **Education.** Distance learning and Internet research are enabled, allowing students anywhere to access resources and obtain realtime instruction from qualified educators

that might not otherwise be available in their local community.

- **Healthcare.** Remote or small clinics can be connected to experts and medical centers throughout the country, broadening access to medical expertise and specialties.
- **Jobs & Productivity.** The availability of broadband access is critical to attracting new businesses and giving existing businesses the ability to compete. With broadband access, worker productivity increases, jobs are created, and wages and the tax base grow.
- **Homeland Security.** Local public safety officials can get timely access to the information they need to assess and act on threats. In times of crisis or a natural disaster, getting accurate information to residents can be a life saver. Informed citizens are better prepared to help themselves and their neighbors in times of need.

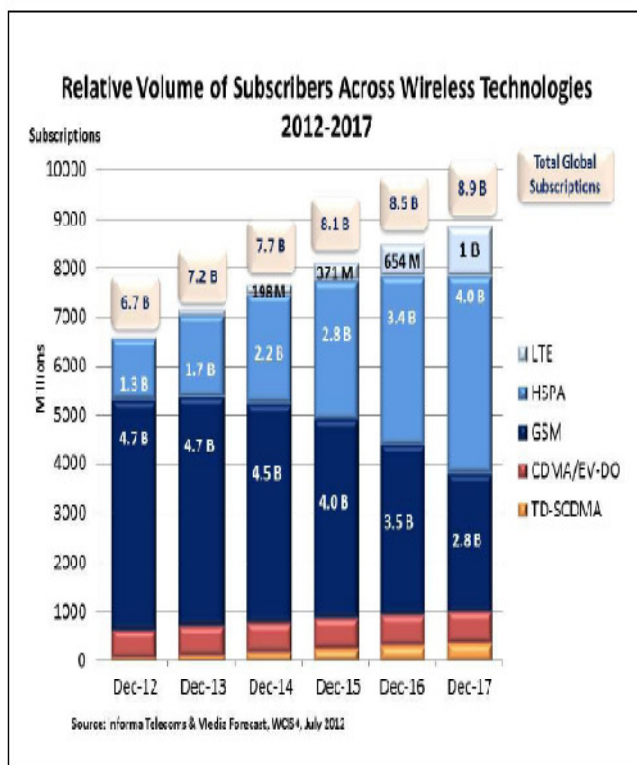


Fig. 2. Relative Volume of Subscribers Across Wireless Technologies

7. HELP BY LOCAL GOVERNMENTS

Local governments can play an active role to bring the advantages of wireless broadband to their citizens.

- **Permits.** In order to deploy a wireless network, operators typically must mount small, safe antennas to towers, buildings, or other tall structures and, in many

cases, on their customers' rooftops. A WISP may need the appropriate permits for the placement of antennas necessary to ensure system coverage. With respect to equipment on the customer's premises, however, federal regulations generally prohibit local jurisdictions from requiring permits.

- **Access to Rights-of-Way and Public Property:** Sometimes local governments control access to the most beneficial structures (such as a water tower or high rooftop) that would enable a WISP to reach a large portion of the community. Some WISPs may also use access to smaller structures in rights-of-way, such as streetlights or telephone poles to set up their networks.
- **Flexibility:** WISPs often vary in size and use different and multiple frequencies. They may have very different requirements and economies than the cellular telephone and other operators with whom local governments may be accustomed to dealing. In order to bring the benefits of broadband to their communities, municipalities may need to work closely with the WISP to ensure that the WISP is able to move forward and that the greatest benefit is brought to the community in terms of service.

8. CONCLUSION

Mobile broadband has become the leading edge in innovation and development for computing, networking, and application development. There are now more smartphones shipped than personal computers. As smartphones and other mobile platforms, such as tablets, increase their penetration levels, they will continue driving explosive growth in data usage, application availability, 3G/4G deployment, and revenue.

The growing success of mobile broadband, however, mandates augmentation of capacity to which the industry has responded by using more efficient technologies, deploying more cell sites, planning for sophisticated heterogeneous networks, and offloading onto either Wi-Fi or femtocells. Some governments that want to lead the mobile broadband technology revolution have responded with ambitious plans to supply more spectrum, while other governments still need to do more by providing more harmonized spectrum soon.

At the end of the Broadband Wireless Era, billions of people worldwide will be communicating wirelessly using devices and services not yet designed. Many of these people will have access to multiple technologies that will allow them choices for an always best-connected advantage. Intel and the members of the Intel Communications Alliance are helping define the Broadband Wireless Era through innovative, wireless-optimized silicon building blocks and platforms, collaboration with other industry leaders on technology and infrastructure design, and the development of new standards.

9. REFERENCE

- [1] A. Yarali, S. Rahman and B. Mbula, "WiMAX: The Innovative Broadband Wireless Access Technology", JOURNAL OF COMMUNICATIONS, VOL. 3, NO. 2, pp: 53-63, APR. 2008
- [2] Brewer E, Demmer M, Du B, Ho M, Kam M, Nedeveschi S, et al. The case for technology in developing regions. Silver Spring, MD: IEEE Computer Society Press; 2005. p. 25–38.
- [3] G. Singh and J. Singh "Comparative Analysis of Broadband Wireless Technologies", International Journal of Computer Science & Technology, IJCST, pp: 69-72 Vol. 2, Issue 4, Oct . - Dec. 2011.
- [4] L. Bai," Analysis of the Market for WiMAX Services", Thesis, Lyngby, Denmark, May 2007
- [5] Mehmet S. Kuran and Tuna Tugcu "A survey on emerging broadband wireless access technologies", Science Direct , Elsevier Science, 2007
- [6] Vinoth Gunasekaran, Fotios C. Harmantzis "Emerging wireless technologies for developing countries", Elsevier, Technology in Society 29 (2007) 23–42
- [7] WiMAX technology overview: (www.intel.com/netcomms/technologies/wimax)
- [8] Rensburg, J.J (2006), "Investigation of the Deployment of 802.11 Wireless Networks", M.Sc Thesis. University of Rhodes: Ghrastown, South Africa.
- [9] Chan, H.A (2005), "Overview of Wireless Data Network Standards and their Implementation Issues", SATNAC Proceedings.
- [10] Intel Corp (2003), "IEEE 802.16 and WiMAX: Broadband Wireless Access for Everyone", [Online] Available http://www.intel.com/ebusiness/pdf/wireless/intel/80216_wimax.pdf
- [11] Black Box (2005), "802.11: Wireless Networking", White Paper, [Online] Available http://www.blackbox.com/Tech_Support/White_Papers/802.11-Wireless-Networking2.pdf
- [12] D.V. Chandra Shekar, V. J. (2005-2008), "Wireless security: A comparative analysis for the next generation networks", Journal of Theoretical and Applied Information Technology.
- [13] White paper, "Understanding WiMAX and 3G for Portable/Mobile Broadband Wireless", Intel, December 2004
- [14] P. Yegani, Cisco Systems white paper, "WiMAX Overview," IETF-64 Nov. 7-11, Vancouver, Canada, 2005, pp. 4.
- [15] Rao B, Parikh MA. Wireless broadband drivers and their social implications. Technol Soc 2003;25:477–89.
- [16] Pentland A, Fletcher R, Hasson A. DakNet: rethinking connectivity in developing nations. Silver Spring, MD: IEEE Computer Society Press; 2004. p. 4–9
- [17] N. Cohen, "What Works: Grameen Telecom's Village Phones," World Resources Inst., 2001, <http://www.digitaldividend.org/pdf/grameen.pdf>.
- [18] M.H. Akhand, "Disaster Management and Cyclone Warning System in Bangladesh," *Early Warning Systems for Natural Disaster Reduction*, J. Zschau and A.N. Koppers, eds., Springer, 2003.
- [19] C. Smith, D. Collins, 3G Wireless Networks, second ed., McGraw-Hill Osborne Media, 2006
- [20] Q. Ni, L. Romdhani, T. Turletti, A survey of QoS enhancements for IEEE 802.11 wireless LAN, Wiley Journal of Wireless Communications and Mobile Computing 4 (5) (2004) 547–566.
- [21] A.S. Tanenbaum, Computer Networks, fourth ed., Prentice Hall, 2003.

Performance Analysis with Multi-antenna for MIMO Wireless System

Monalisa Bhol¹, Sikha Mishra², Mihir Narayan Mohanty³

MIEEE

Department of Electronics & Communication Engineering,
ITER, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha

¹monalisa28bhol@gmail.com, ²mishra.sikha@rediffmail.com, ³mihir.n.mohanty@gmail.com

Abstract: The demand for high data rates increases day by day, but the limited available bandwidth motivates the investigation and new area of research in wireless systems. Increasing demand for higher wireless system capacity has catalysed several transmission techniques, among which is the multiple-input/multiple-output (MIMO) technology has a great attraction and the great part of recent attention. In this paper, it has been compared among various multi-antennas (MA) at both the transmitter and receiver ends for significant capacity achievement. The use of antennas at both sides of the wireless communication link can result in high channel capacity provided the propagation medium is rich. Rayleigh fading has been considered as the propagation medium for verification. Also the performance has been measured in terms of bit error rate (BER) along with the capacity measurement. The result shows its performance.

Keywords: MIMO, Channel Capacity, BER, multi-antenna, Diversity.

1. INTRODUCTION

Wireless systems continue to strive for higher data rates.

This goal is particularly challenging for systems that are power, bandwidth, and complexity limited. However, another domain can be exploited to significantly increase channel capacity by the use of multiple transmit and receive antennas. The employment of multiple antennas at both the transmitter and receiver, known as Multiple Input Multiple Output (MIMO) technologies, enables to greatly improve the link reliability and increase the overall system capacity. It is one of the smart antenna technologies available in several forms. MIMO is a method of transmitting and receiving two or more unique data streams through a single radio channel [1]. MIMO systems are equipped with multiple antennas, at both the transmitter and receiver in order to improve communication performance.

MIMO technologies overcome the deficiencies of the traditional methods through the use of spatial diversity. Data in a MIMO system is transmitted over M transmit antennas to N receive antennas supported by the receiver terminal. MIMO systems are used in wireless communication for

enhancement of capacity and BER. Diversity gain and spatial multiplexing gain are the two main advantages of MIMO systems that are used to study the effect of increase in bit rate with increasing the number of transmitter and receiver antennas [2]. In this paper we have observed the variation in the capacity of MIMO system with the number of transmitting antennas, receiving antennas as well as by increasing the SNR. It is of great interest to characterize and model the MIMO channel for different conditions in order to predict, simulate, and design high performance communication systems. The simulation results shows that the performance of the MIMO system improves with the number of transmit and receive antennas in terms of capacity and bit error rate (BER).

The paper is organized as follows. Section-II contains the related literature; section-III describes the model approach and the performance measure of this work. Section-IV shows the result and finally section-V concludes the work along with the future direction.

2. RELATED LITERATURE

In recent years, a lot of attention has been drawn to systems with multiple element transmitter and receiver arrays, because they can achieve very high spectral efficiencies. The spectral efficiency that can be exploited in MIMO systems depends on a number of phenomena, including the average received power of the desired signal, thermal and implementation-related noise, as well as co-channel interference [3]. In [4], authors investigate the effect of Rician factor (K) and the correlation coefficient (r) on the capacity and diversity of multi-input multi-output (MIMO) systems. They have presented the view point that the loss or gain in the capacity or diversity can be considered as an equivalent gain in the SNR. In [5], authors investigate the capacity distribution of spatially correlated, multiple-input-multiple-output (MIMO) channels. In particular, authors derive a concise closed-form expression for the characteristic function (c.f.) of MIMO system capacity with arbitrary correlation among the transmitting antennas or among the receiving antennas in frequency-flat Rayleigh-

fading environments. Using the exact expression of the c.f., the probability density function (pdf) and the cumulative distribution function (CDF) can be easily obtained, thus enabling the exact evaluation of the outage and mean capacity of spatially correlated MIMO channels. In [6], the paper reviews recent research findings concerning antennas and propagation in MIMO systems. Issues considered include channel capacity computation, channel measurement and modelling approaches, and the impact of antenna element properties and array configuration on system performance. Throughout the discussion, outstanding research questions in these areas are highlighted.

As the user's needs for higher data rates grow and bandwidth is becoming an expensive commodity, MIMO systems have become an especially attractive potential solution for wireless applications that are inherently power and complexity limited [7]. The multiple antennas in MIMO systems can be exploited in two different ways. One is the creation of a highly effective antenna diversity system; the other is the use of the multiple antennas for the transmission of several parallel data streams to increase the capacity of the system [8].

Antenna diversity is used in wireless systems to combat the effects of fading. If multiple independent copies of the same signal are available, we can combine them to a total signal with high quality, even if some of the copies exhibit low quality. For spatial MIMO configurations, all the sub channels of H are identically distributed. MIMO architecture has the potential to dramatically improve the performance of wireless systems. The increasing demand for capacity in wireless systems has motivated considerable research aimed at achieving higher throughput on a given bandwidth [9].

3. PROPOSED MODEL

In MIMO systems, the propagation of electromagnetic waves from a transmitter to a receiver is characterized by the presence of multi-paths due to various phenomena such as reflection, refraction, scattering and diffraction and the performance of MIMO systems is largely dependent on the propagation medium.

By applying MIMO technology, we can directly take the advantage of two important properties, i.e., Diversity and Multiplexing. Diversity indicates the replicas of the signal, bearing the same information at the receiving end and Multiplexing is a transmission technique in MIMO wireless communication to transmit independent and separately encoded data signals, so-called streams from each of the multiple transmit antennas. Therefore, the space dimension is reused, or multiplexed, more than one time.

The simplified model for the MIMO system can be characterized by its impulse function. The input-output

relation using the transmitted and received signal along with the channel noise can be represented by equation (1).

The elements of narrowband MIMO channel matrix are assumed to be independent and identically distributed (i.i.d.) to study the MIMO channel capacity. In reality, however, due to insufficient spacing between antenna elements and limited scattering in the environment, the fading is not always independent causing a lower MIMO channel capacity compared to the ideal, i.i.d. case. Therefore the proposed MIMO channel models should take this effect into account.

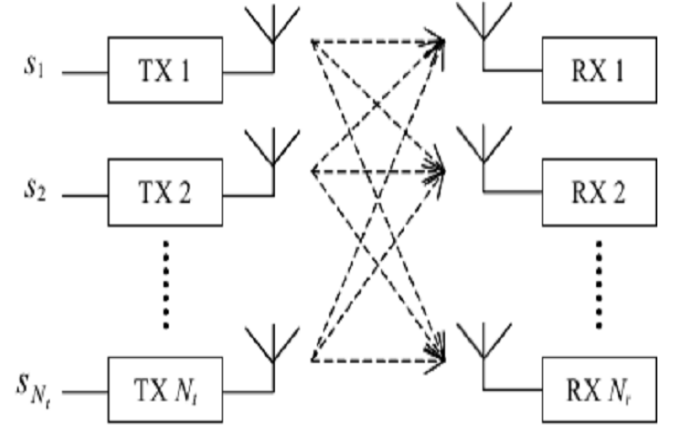


Fig. 1. MIMO system with m transmit and n receive antenna elements.

$$y^{(f)} = H^{(f)}x^{(f)} + n^{(f)} \quad (1)$$

where,

$x^{(f)} = [x_1^{(f)}, \dots, x_N^{(f)}]$, transmitted signal

$y^{(f)} = [y_1^{(f)}, \dots, y_N^{(f)}]$, received signal

$n^{(f)} = [n_1^{(f)}, \dots, n_N^{(f)}]$, is additive white Gaussian noise (AWGN).

$H^{(f)}$ = Channel impulse response matrix

The assumptions have been made for evaluation is as follows:

- 1) The fading at the different antenna elements is assumed to be i.i.d Rayleigh fading. This is fulfilled if the directions of the multipath components at the transmitter and receiver are approximately uniform. Also the antenna elements are spaced far apart from each other.
- 2) The flat fading can be satisfied provided that the coherence bandwidth of the channel is significantly larger than the transmission bandwidth.
- 3) It has been assumed that the receiver has perfect knowledge of the channel. For the transmitter, we will

analyse both cases where the transmitter has no channel knowledge, and where it has perfect channel knowledge.

- 4) When talking about capacity, we also assume that the channel is quasi-static. By quasi-static, we mean that the coherence time of the channel is so long that “almost infinitely” many bits can be transmitted within this time. Thus, each channel realization is associated with a (Shannon - AWGN) capacity value. The capacity thus becomes a random variable, described by its cumulative distribution function (cdf).

A. Channel Capacity and BER

It is well known from Shannon's theorem that a particular SNR can give only a fixed maximum capacity [10]. In the system it has been considered that for the fading channel, the SNR constantly changes. As the rate of fade changes, the capacity changes with it. The capacity is given by the formula:

$$C = B \log_2(1 + SNR) \quad (3)$$

For a MIMO system the channel capacity is given by the formula:

$$C = M \log_2 \det(I + SNR) \quad (4)$$

where,

C = channel capacity

$M = M$ is the minimum of N_T and N_R

I = identity matrix

N_T = number of transmitting antennas

N_R = number of receiving antennas

The channel capacity gradually increases, with increase in number of transmitting and receiving antennas,

The BER is an approximate estimation of the bit error probability. The number of bit errors is the erroneous number of received bits of a data stream over a communication channel. The error may cause due to noise, interference, distortion or bit synchronization. BER for QPSK modulation is given by:

$$BER = \frac{1}{2} \operatorname{erfc} \sqrt{\left(\frac{2E_b}{N_o}\right)} \quad (5)$$

where,

E_b = Energy per bit

N_o = Noise spectral density

Result and Discussion

In Fig. 2, shows the performance of capacity in the wireless channel. The result is measured according to the various SNR values. The capacity of the MIMO channel has been

simulated for number of transmitter and receiver antennas, such as 2×2 , 3×3 , 4×4 , and 8×8 MIMO systems. It is observed that capacity gradually increases with the number of antennas. The proposed analysis allows evaluation of the capacity and the outage capacity for MIMO systems. The system capacity with respect to the outage probability has been shown in Fig. 3. It is noticed that low probability means low capacity. We can increase the number of antennas to increase capacity for a given outage probability. As BER is one of the performance parameters, it has been verified with various modulation techniques such as BPSK, QPSK, and QAM. BER performance for QPSK has been shown in Fig. 4, i.e., BER versus SNR, with different number of transmitting and receiving antennas (2×2 , 3×3 , 4×4 , and 8×8).

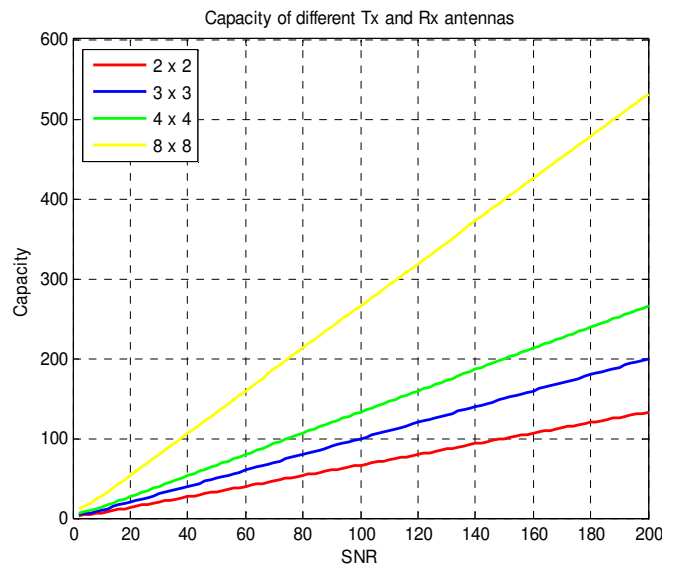


Fig. 2 Performance of capacity with respect to SNR.

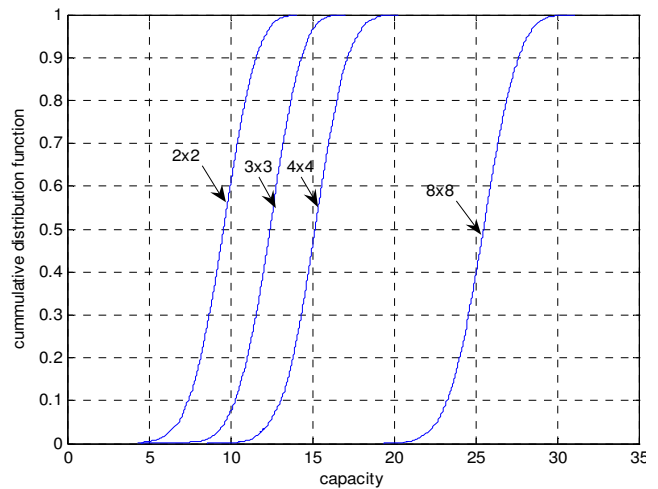


Fig. 3 System capacity as a function of outage probability.

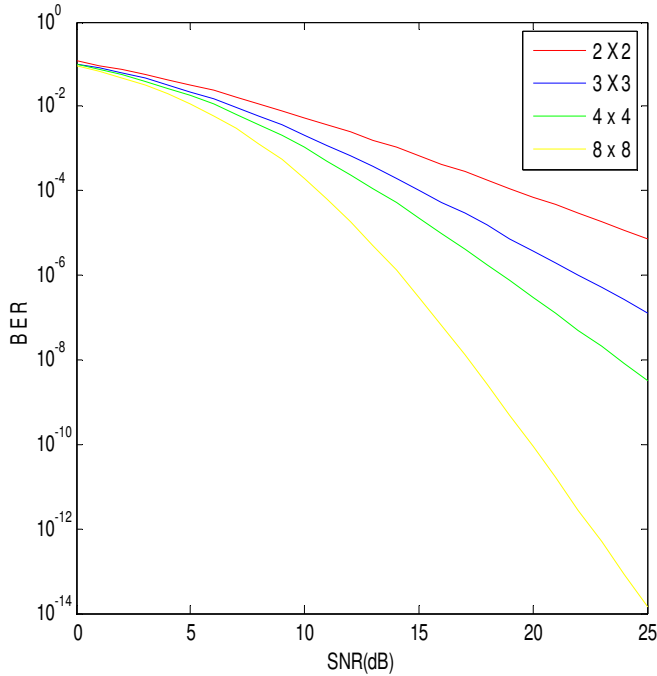


Fig . 4. Performance of BER.

4. CONCLUSION

MIMO systems with reduced complexity are now being used for third-generation cellular systems (W-CDMA), and are discussed for future high-performance mode of wireless networks. The multiple antennas in MIMO systems can be exploited in two different ways. One is the creation of a highly effective antenna diversity system; the other is the use of the multiple antennas for the transmission of several parallel data streams to increase the capacity and increase the BER performance of the system. Multiple antenna communications technologies offer significant advantages over single antenna systems. These advantages include extended range, improved reliability in fading environments and higher data throughputs.

REFERENCES

- [1] Akhilesh Kumar, Anil Chaudhary: "Channel Capacity Enhancement of Wireless Communication using MIMO Technology", International Journal of Scientific & Technology Research Volume 1, Issue 2, March 2012.
- [2] Mahesh Kusuma, Mohan Amgothu, Bhadrn Amgothu, "CAPACITY ANALYSIS OF MIMO SYSTEMS", International journal of Advances in Computer Networks and its Security.
- [3] Deeparani Mishra, Sikha Mishra, Mihir N. Mohanty, "Estimation of MIMO-OFDM Based Channel for High Data Rate Wireless Communication", IJCSIT, Vol. 2 (3) , 2011, pp.1263-1266.
- [4] Syed M. Tabish Qaseem, and Adel A. Ali, Effect of Antenna Correlation and Rician Fading on Capacity and Diversity Gains of Wireless MIMO Channels, International Symposium on Wireless Communications (ISWSN'05) 2005.
- [5] Marco Chiani, Moe Z. Win, and Alberto Zanella, On the Capacity of Spatially Correlated MIMO Rayleigh-Fading Channels, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 49, NO. 10, pp. 2363-2371, OCTOBER 2003.
- [6] Michael A. Jensen, and Jon W. Wallace, A Review of Antennas and Propagation for MIMO Wireless Communications, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 52, NO. 11, pp. 2810-2824, NOVEMBER 2004.
- [7] Persefoni Kyritsi, Donald C. Cox, Reinaldo A. Valenzuela, Peter W. Wolniansky, "Correlation Analysis Based on MIMO Channel Measurements in an Indoor Environment", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 21, NO. 5, JUNE 2003.
- [8] Andreas F. Molisch, Moe Z. Win, "MIMO Systems with Antenna Selection – an Overview", First printing, TR-2004-014, March 2004.
- [9] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.
- [10] J. G. Proakis, 'Digital Communications', 4th edition, McGraw-Hill, New-York, 2000.

Industrial Ethernet: Convergence of Communication and Information Technology with Control Systems

H.K. Gopinath¹, Ajeet S Rawat², Mukesh G³, Rakshit G.⁴

*Rajasthan Institute of Engineering Technology, Jaipur, India
gopisigs87@gmail.com*

Abstract: The manufacturers of various industries face tremendous challenges with the constant demand to increase productivity and reduce operating costs. As most manufacturer grow from national to international and final endeavor to a global player, introduction of technological advances helps rapid growth. A most viable and powerful industrial tool being implemented today to streamline production and leverage technology is by migrating to Industrial Ethernet. This involves integrating with reliable, enterprise-wide connectivity using Information Technology (IT) upto the floor level and with use of Internet, and meshed with High Technology Communications using Wired and Wireless Technology. This provides the highest level of visibility, control and flexibility to enhance production. To accommodate evolving networking requirements, network protocols integrate with industrial equipment and control systems to communicate crucial data with precision end equipments like robots and similar equipments. Integration of Industrial Control Systems and Communication has never seen such a synergy as part of a continuing effort to make Industrial organizations more efficient and flexible. Manufacturers are rapidly migrating to Industrial Ethernet technology to network their industrial automation and control systems thereby interconnect plant and business systems to achieve business objectives and improved performance Index. Industrial Ethernet enables a more flexible, responsive system that encompasses real-time data from the production floor. This end-to-end networking architecture provides connectivity, collaboration, and integration from the device level to enterprise business systems. This review paper provides overview of Industrial Ethernet Technology, the requirements and benefits in industrial networking environments, which is also called the Industrial Ethernet..

Index Terms: Industrial Ethernet, Protocols, Fieldbus, OSI Model, Quality of service (QoS).

1. INTRODUCTION

Used to connect various machines, devices and office equipment, networks enable manufacturers to visually manage all production parameters for continuous and complete control over data using varied Information Technology, thus streamlining efficiency and minimizing production downtime. By understanding the need for enterprise connectivity, the various challenges and considerations associated with implementing network protocols, users can maximize data acquisition and management capabilities. Industrial Ethernet can be

explained with an illustration of activities in an soft drink bottling plant. The filling operation can be assumed to be run by an Industrial Ethernet network. The network works well because it uses “electronic handshaking” to ensure message delivery. Say the bottling device begins filling a bottle at the command of the controlling Programmable Logic Controller (PLC) the chip controlling all activity. The PLC is also responsible for sending the “stop filling” command when the bottle is full. If the message is lost on the network, the PLC is aware because it does not receive a delivery response, (part of the handshaking) so it knows to resend the command. Where as in the office setting, such a lost transmission is rarely important. If a web page gets lost in transmission, the user simply presses “refresh.” In the production setting, though, we cannot wait for the soft drink to spill on the floor before someone manually turns off the filler. The handshake saves the plant from spillover of soft drink, saves money and time.

In an industrial Ethernet network, we also incorporate collision detection. If two messages collide in our network, the controlling PLC can resend the message to the device until it receives a ack notice for the device. The plant continues its controlled pour, and no one is crying over wasted drink. Such plants require a few dozen bottlers, valves, sensors, and a PLC in this network.

The operation must run at peak efficiency and an usual office Ethernet network would not accomplish this goal. Industrial units maintain networks to support their factory floor operations and corporate business operations separately. While the corporate IT network supports traditional Administrative functions of the organisation, the control-level network connects control and monitoring devices and lastly the device-level network links the controllers with the plant floor’s I/O devices. Instead of using separate networks in different planes, a systematic integration of all these networks leads to cost cut, more efficiency, better control and monitoring and enhanced performance. Industrial Ethernet has the capability to integrate Company’s administrative, control level and device-level networks in a single network infrastructure. The concept is given in the figure 1.

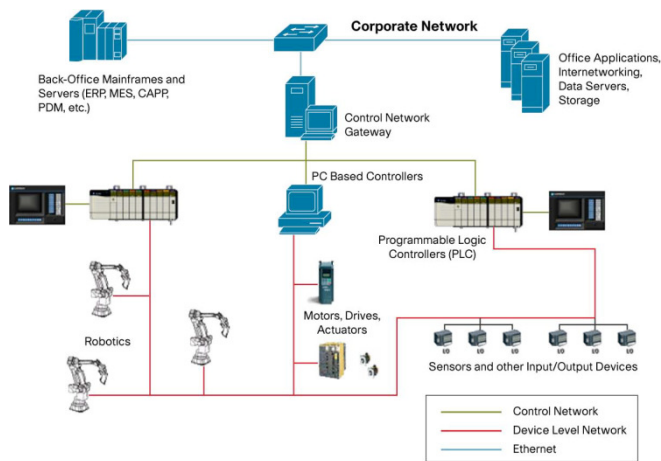


Fig. 1. Integration of Management level, control and Device level network

2. ANALOGY WITH OFFICE ETHERNET

To understand better we draw the analogy of industrial Ethernet with a office automation Ethernet with which we are familiar.

A Operations concerns

An important area of concerns the cost of downtime. When a network goes down in your office setting, it is an inconvenience, and some work may be impossible. Often though, an employee will simply need to move on to another task and tackle it without use of the Internet. Where as in a production setting, that downtime is costly. Assembly lines operating with continual processes can be rendered nonfunctional if one aspect fails. Critical processes could be ruined, leading to lost of material and money. Imagine a factory producing tempered glass for windows. A continuous flow of glass moves from pour, to cut, over an assembly line a mile long. The glass flow progresses through specific heat-ups, cool-downs, and rests to properly temper it to meet production specifications. If the line suddenly seized, the factory would be left with a mile of scrap glass. Much of it that would need to be removed manually due to the fact it had cooled hard on a portion of the line that was meant to deal with hot malleable glass.

B Security

In an office setting, the information traveling through the network can be confidential and important, thus an office Ethernet network must guard against unauthorized use. The same is true in an industrial application. Another security threat in the industrial setting is the risk that an employee may break the system accidentally, creating a Garbage In/Garbage Out scenario or bringing the device or network to a complete halt.

C Differences between an Office and shop floor

You would not expect to see someone in the industrial setting wearing formal suits or expensive leather shoes because it is much more suitable for them to be wearing jeans and steel-toed boots. These choices offer more protection from the environmental factors in the factory. The same attire considerations need to be taken for Ethernet networks. Industrial Ethernet cables, switches, and connectors need to withstand the unique and harsh criteria in an industrial setting.

D Environmental concerns

Heat and cold are two factors that can have a major effect on a network. Cold is particularly damaging, at relatively cold levels, near freezing, a cable is susceptible to impact, which can cause a break in the cable, destruction of the protective jacket, or attenuation. At even colder temperatures, the cable may become brittle and break through no large force, but instead through simple bending. Heat is also damaging. The protective jacket may melt, leading to shorts and vulnerability. Heat also causes attenuation over time.

Chemicals may cause a jacket to dissolve or change shape, leading to a shorter life and worse performance. Some solvents can also directly impact the internal cable should the protective jacket not be effective. Radiation, especially UV Radiation from sunlight, can cause discoloration and degradation of the jacket. Humidity can also degrade the cable. The industrial Ethernet environment is harsh, and office Ethernet applications are not created for such environments. Taking measures to physically protect cables and connectors can minimize, or even negate, the effects of an industrial environment.

Electric and magnetic noise generated by large motors and high voltage devices can distort data transfers on the network. Some processes may create vibration, which can cause degradation of the jacket and disconnection, if poor connectors are used. Hence, when designing an industrial Ethernet network, one must consider options that make network reliable.

3. INTEGRATION OF ETHERNET ON INDUSTRIAL PLATFORM

Industrial Ethernet networks that use intelligent switching technology can offer a variety of advantages compared to traditional industrial networks. The technology can be deployed using a switched Ethernet architecture and has proven successful in multiple critical applications in different markets. Because the technology is based on industry standards, Industrial Ethernet enables organizations to save money by moving away from expensive, closed, factory floor-optimized networks. Using standard Ethernet

technologies also reduces overall risk and provides investment protection, as manufacturers and automation vendors can take advantage of continued industry investment and innovation in compatible technologies (such as video and voice over IP). By providing a scalable platform that can accommodate multiple applications, Ethernet-based automation systems can increase flexibility and accelerate deployment of new applications in the future. At the same time, Ethernet delivers the network security, performance, and availability required to support critical manufacturing applications. To deploy this technology, engineers on the manufacturing floor should be familiar with some of the important concepts behind Industrial Ethernet. Once considered a solution that was limited to corporate network environments, Ethernet technology has proven to be a robust alternative that can meet the unique needs of the manufacturing environment.

On the OSI Model the Ethernet reside in the layer -2 i.e the data link layer and performs same task as done by some of the fieldbus. Layer 3 takes care of the logical addressing and routing (which way to send data). Its most common implementation uses the Internet Protocol (IP), which is the core of World Wide Web addressing and routing. Layer 4, the last of the lower layers, is the transport layer. It ensures that data is delivered error-free and in the correct sequence. Industrial Ethernet is broader than traditional Ethernet technology. While Ethernet technology refers only to Layers 1 and 2, most Industrial Ethernet solutions also encompass Layers 3 and 4, using IP addressing in Layer 3, and Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) in Layer 4, in what is referred to as the IP suite. The upper layers of the OSI reference model are responsible for application tasks and are usually implemented only in software. The highest layer, the application layer, is closest to the end user. Both users and application-layer processes interact with software applications that involve network communications. Figure 2 explains pictorially.

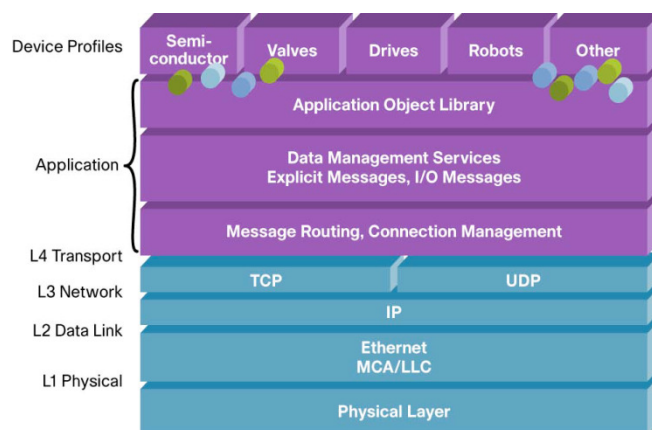


Fig. 2. Industrial Ethernet with OSI Model

A Brief overview of Ethernet?

Ethernet is by far the most widely used LAN technology today, connecting the world's LAN connected PCs and workstations. Ethernet refers to the family of computer networking technologies covered by the IEEE 802.3 standard, and can run over both optical fiber and twisted-pair cables. Over the years, Ethernet has steadily evolved to provide additional performance and network intelligence. This continual improvement has made Ethernet an excellent solution for industrial applications. Today, the technology of Wi-Fi reaches the 1Gbps mark, researchers at standards body IEEE as on Aug 2012 have planned to boost wired connections to 1000 times the speed of their wireless counterparts. As the core networking usage is doubling every 18 months on average, IEEE expects the amount of bandwidth required to raise 100-fold every 10 years. The Higher-Speed Ethernet Consensus group, the team responsible for defining the standard, is debating two options at present, one which wants the next-generation standard to max out at 400Gbps, and another that is aiming for an impossibly-fast 1Tbps connection. However the basic usable Five data rates are as follows

- 10BASE-T Ethernet delivers performance of up to 10 Mbps over twisted-pair copper cable.
- Fast Ethernet delivers a speed increase of 10 times to 100 Mbps while retaining many of Ethernet's technical specifications.
- Gigabit Ethernet extends the Ethernet protocol even further, increasing speed tenfold over Fast Ethernet to 1000 Mbps, or 1 Gbps.
- 10 Gigabit Ethernet, ratified as a standard is an even faster version of Ethernet. Its high data rate of 10 Gbps makes it a good solution to deliver high bandwidth in WANs and metropolitan-area networks (MANs).
- Wi Fi to 1GBPS and wired to 1Tbps is planned by 2014

B What Is Industrial Ethernet?

Industrial Ethernet (IE) applies the Ethernet and Internet Protocol (IP) suite of standards developed for data communication to manufacturing control networks. It is based on the IEEE 802.3 standard, which is similar to Ethernet found in an office environment but adapted to an industrial environment for automation. Recognizing that Ethernet is the leading networking solution, many industry organizations are porting the traditional fieldbus architectures to Industrial Ethernet. Industrial Ethernet applies the Ethernet standards developed for data communication to manufacturing control networks. Using IEEE standards-based equipment, organizations can migrate all or part of their factory operations to an Ethernet environment at the pace they wish. For example, Common

Industrial Protocol (CIP) has implementations based upon Ethernet and the IP protocol. Most controllers (with appropriate network connections) can transfer data from one network type to the other, leveraging existing installations, yet taking advantage of Ethernet. The advantage of Industrial Ethernet is that organizations and devices can continue using their traditional tools and applications running over a much more efficient networking infrastructure.

Industrial Ethernet not only gives manufacturing devices a much faster way to communicate, but also gives the users better connectivity and transparency, enabling users to connect to the devices they want without requiring separate gateways.

C Technology Tailored for Manufacturing

Although Industrial Ethernet is based on the same industry standards as traditional Ethernet technology, the implementation of the two solutions is not always identical. The primary difference between Industrial Ethernet and traditional Ethernet is the type of hardware used. Industrial Ethernet equipment is designed to operate in harsh environments. It includes industrial-grade components, convection cooling, and relay output signaling. And it is designed to operate at extreme temperatures and under extreme vibration and shock (and other conditions). Power requirements for industrial environments differ from data networks, so the equipment runs using 24 volts of DC power. To maximize network availability, Industrial Ethernet equipment includes fault-tolerant features such as redundant power supplies, modular in order to meet the highly varying requirements of a factory floor.

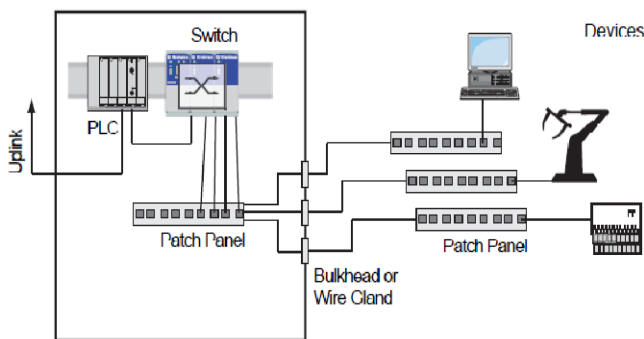


Fig. 3. Typical industrial topology

The Industrial Ethernet automation and control protocols, differ significantly, from standard Ethernet implementations. In most automation and control applications, 80 percent of the network traffic is local, one local device talking to another local device often using multicast (one sender, many receivers) packets. In most IT installations, the reverse is true where 80 percent of the network traffic is routed to

external locations (such as the data center or the Internet) using unicast (one sender, one receiver) packets. Automation and control systems also differ from other applications in their need for determinism and real-time network requirements, quick and consistent transmission of the data. Ethernet and the IP protocol suite have developed a number of technologies and features that support these requirements. Figure 3 below shows a typical industrial topology of use of Industrial Ethernet.

Managed switches provide performance management, diagnostics, and security capabilities that are not supported on unmanaged switches. These types of features allow the network administrator to configure the switch to provide host of issues, some of the most important features on intelligent managed switches in an industrial environment include

- Intelligent switching platforms can dynamically configure the interfaces so that traffic is forwarded only to ports associated with requested data. This feature reduces the load of traffic crossing the network and relieves the client devices from processing unneeded frames.
- Network analyzers allow traffic analyzers to remotely monitor any port in a network (also known as port mirroring), which saves organizations time and money and reduces the amount of hardware that must be deployed to monitor and optimize network usage.
- Managed switches play a major role in security approach, as the first point of access to a network and system. It starts with port security and settings that control which devices can connect. And managed switches can be configured to reduce or eliminate common types of attacks intentional or unintentional.
- Diagnostics is a critical factor to have the right information, and these provide a host of diagnostic information helpful to resolve network and device issues occurring in the automation and control systems. Critical diagnostic information directly manages and control the automation and control applications, like any other device on the factory floor. This enables production personnel to monitor and maintain the network.

When implementing an Industrial Ethernet solution, it is important to select Ethernet products that offer the intelligent features required to support manufacturing applications. The important qualities behind an intelligent Industrial Ethernet solution include network security, reliability, manageability (ease of use) and achieving determinism. (Achieving determinism implies handling time-critical cyclic data across the network.). One such proprietary protocol used in the Industrial Ethernet is the Resilient

Ethernet Protocol (REP) of Cisco technology that is designed to achieve fast convergence in a ring topology or a variety of complex topologies. REP is a segment technology and can interoperate with other protocols.

Quality of Service implies the network must be able to distinguish and give priority to different types of traffic. By giving priority to different types of traffic, the network can deliver real-time network services, low latency and jitter and minimal packet loss when the network infrastructure is under optimum load. Figure 4 shows One such process control system involving Two plants located at Two different places and connected on Optical Fiber as a core backbone network.

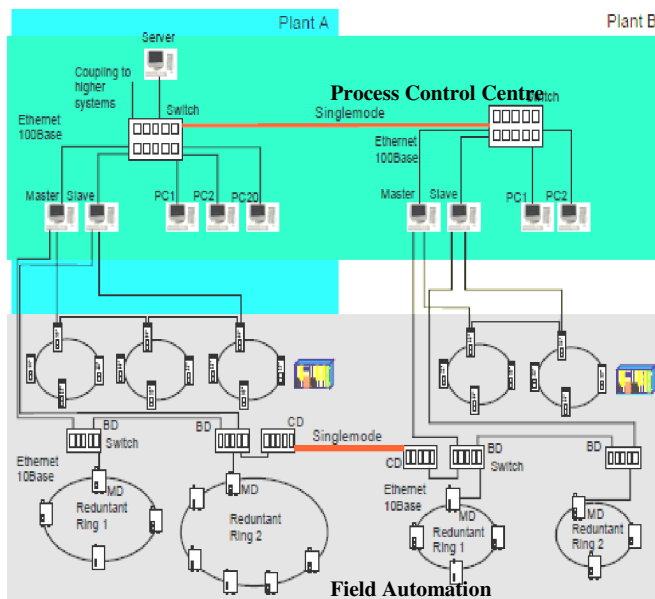


Fig. 4. Connectivity between Two Plants

4. BENEFITS OF INDUSTRIAL ETHERNET

Industrial Ethernet switches (the core building block of industrial networks) are optimised for harsh environments

- Industrial switches have hardened enclosures, dual power inputs, shock and vibration approvals, direct DC power, DIN-rail mounting capability and high MTBFs to provide a more suitable solution for industry.
- Using Ethernet instead of serial (RS232, or RS422) communications massively increases the flexibility of the installed network. A single cable can be used for video, voice and many different types of data. Traditional fieldbus systems were limited to one task with no scope for diversification.
- Power over Ethernet (PoE) allows power to be delivered over the same piece of Cat 5 or Cat 6 cable that

transmits Ethernet data. This means that devices such as IP cameras, gas analysers and embedded computers can be conveniently located without the cost of installing an additional power spur.

- Fast network redundancy is a key driver for adopting Industrial Ethernet. It is possible to build a network of devices that can continue to communicate with each other even if a cable is broken or unplugged or if one of the network switches fails. Many safety critical systems must exhibit high reliability and fault tolerance.
- Most Industrial Ethernet hardware has a wide temperature range - typically -40° to $+70^{\circ}\text{C}$ and conformal coating (a protective coating for the PCB) allowing the switches to be used outdoors.
- Most applications that use Ethernet use TCP/IP hence it is possible to connect industrial networks to each other allowing data sharing and remote control even if sites are geographically separated..
- Another benefit that arises from Ethernet's close affiliation with IP is the ability to use mobile phone networks to communicate with industrial equipment. 2.5G and 3G support IP based which means a PLC in a vehicle that is travelling around in Connaught place in New Delhi can stay in permanent contact with a PC or server located at Gurgaon factory floor several miles away. Using cellular communications provides a data cable the length and breadth of the globe and facilitates applications that were never previously possible.
- Ethernet consists of a set of well defined standards and protocols that encompass many different media types. Fibre optic communication is especially useful in industrial applications for two of its key features like Electromagnetic noise immunity, and inherent advantage over distance. Some of the popular protocols like EtherNet/IP, Modbus TCP and PROFINET assists manufacturers in selecting the ideal networking solution for their critical communication needs in this field of Industrial Ethernet.

A Impact of Costs on Network Failure

If a commercial Ethernet switch or cabling system fails in an industrial environment, the real cost to the manufacturer is typically much more than just the replacement cost of the components. In fact, the cost of the parts themselves are typically only a fraction of the cost involved. The real cost may be much broader in Industry with different problems, like

- Downtime in an automotive assembly plant capable of producing one vehicle per minute could equate to a One to Two lacs per minute loss of profits for small car production and much more for SUV and large truck production,

- A greater need for repetitive repairs, if the cable or switch performance is intermittent.
- A loss of worker or environmental safety.

What costs would company incur, in terms of liability, if a poorly chosen switch or cable fails in a safety-critical application. In addition, the cables in an industrial environment are more likely to experience pulling forces far beyond those of the initial installation process as equipment is rearranged on the plant floor. ARC Advisory Group the worldwide market for Industrial Ethernet (leading technology research and advisory firm for industry and infrastructure) expects industrial Ethernet to proliferate at a Compounded Annual Growth Rate (CAGR) of 51.4 percent over the next five years

B. Wireless Technology In Industrial Ethernet

Wireless Technology is advancing at a high rate, making a wide range of applications possible including data networking, wireless sensors, wireless actuators and wireless controllers. Wireless standards for automation and control are also developing around various wireless technologies. Industrial Ethernet networks and protocols have become a workhorse of automation systems. Wireless technology facilitates to communicate with sensors embedded or placed at locations difficult to access especially in special devices. Ethernet is not ideal in all applications with some caveats like

- COTS using Ethernet is implemented with widely adopted standards for hardware, software, core protocols, cabling, and connectors.
- Interfacing to enterprise systems is much easier when automation systems use the same Internet Protocol (IP) structure used by information systems. This has been a big plus for the industry.
- Ethernet provides a big pipe for high throughput, efficiently communicating a great deal of information between controllers, and rest of the systems to support modern manufacturing systems.
- The spoke and hub wiring of Ethernet is fine for controllers but creates a great deal of cost for sensor and actuator connections

- Cyber security is becoming the big concern that is likely to keep growing.

C. Use of industrial Ethernet in India.

Though India is a large market for use of Industrial Ethernet, the manufacturer are realizing the importance and accrued benefits in adopting this technology. Top vehicle manufacturer with collaborations are to some extent using industrial Ethernet applications. China is a market where no automation supplier has achieved a commanding lead in either the process or factory automation market.

5. CONCLUSION

The performance demands and characteristics of automation and control networks are fundamentally different from those of the commercial world where Ethernet technology originated. Reliability, bandwidth, and determinism continue to be the critical performance concerns for industrial users, and a number technology, standards, and practices have proven effective at meeting each of these demands. As industrial Ethernet manufacturers and users both continue to develop greater experience and expertise, there is no doubt that even greater heights of performance will be reached.

REFERENCES

- [1] Ronald Dietrich, Industrial Ethernet. from office to machine, Harting Group, 2005, pp17-19,25-49.
- [2] Security for industrial automation, White paper by ABB,2010, pp-06-10.
- [3] Industrial Ethernet for control and automation, Product brochure,2011, MOXA.
- [4] Industrial Ethernet a control engineer guide,White paper, 2007,CISCO.
- [5] Belden, Industrial Ethernet user guide, White paper, HIRSCHMANN.
- [6] Martin Rostan, Industrial Ethernet technology devices, White paper, 2011, EtherCAT Technology Group .
- [7] Evaluating Ethernet Intelligence,White paper,Turck Works
- [8] Ten key benefits of industrial Ethernet, Technical paper, Amplicon
- [9] John Rinald,Industrial Ethernet rage, 2010 edition of wireless and Ethernet, MAVERICK Technologies

Security Management on Telecommunication Network

Abha Jaiswal¹, Ravi Prakash Jaiswal²

(Research Scholar), UPRTOU, Allahabad, India

abha_j27@yahoo.co.in

(Senior Technical Assistant), MGKV, Varanasi, India

ravijaiswal@rediffmail.com

Abstract: Telecommunication networks are today an inseparable part of social interaction and critical national infrastructure. Protecting these networks from malicious attacks, that could lead to unavailability or loss of integrity and confidentiality of network services, is thus an important aspect that cannot be ignored. An effective and robust security programme should be implemented to protect telecommunication networks from such attacks. This paper presents some of the important security challenges to current telecommunication networks, the need for security management and an approach to manage security for these networks.

Keywords: tele communication; Networking; data security; data communication Telecommunication Networks, Security Management, Network Security;

1. INTRODUCTION TO TELECOMMUNICATION NETWORK

A **telecommunications network** is a collection of terminals, links and nodes which connect to enable telecommunication between users of the terminals. Each terminal in the network have a unique address so messages or connections can be routed to the correct recipients. The collection of addresses in the network is called the address space.

The links connect the nodes together and are themselves built upon an underlying transmission network which physically pushes the message across the link, using circuit switched, message switched or packet switched routing.

Telecommunications networks are:

- computer networks
- the Internet
- the telephone network
- the global Telex network

- the aeronautical ACARS network

A. Messages and protocols



Fig.. 1. Example of how nodes may be interconnected with links to form a telecommunications network

Messages are generated by a sending terminal, then pass through the network of links and nodes until they arrive at the destination terminal. It is the job of the intermediate nodes to handle the messages and route them down the correct link toward their final destination.

These messages consist of control (or signaling) and bearer parts which can be sent together or separately. The bearer part is the actual content that the user wishes to transmit (e.g. some encoded speech, or an email) whereas the control part instructs the nodes where and possibly how the message should be routed through the network. A large number of protocols have been developed over the years to specify how each different type of telecommunication network should handle the control and bearer messages to achieve this efficiently

Components

All telecommunication networks are made up of five basic components that are present in each network environment

regardless of type or use. These basic components include terminals, telecommunications processors, telecommunications channels, computers, and telecommunications control software.

Terminals are the starting and stopping points in any telecommunication network environment. Any input or output device that is used to transmit or receive data can be classified as a terminal component.^[1]

Telecommunications processors support data transmission and reception between terminals and computers by providing a variety of control and support functions. (i.e. convert data from digital to analog and back)^[1]

Telecommunications channels are the way by which data is transmitted and received. Telecommunication channels are created through a variety of media of which the most popular include copper wires and coaxial cables (structured cabling). Fiber-optic cables are increasingly used to bring faster and more robust connections to businesses and homes.^[1]

In a telecommunication environment computers are connected through media to perform their communication assignments.^[1]

Telecommunications control software is present on all networked computers and is responsible for controlling network activities and functionality.^[1]

Early networks were built without computers, but late in the 20th century their switching centers were computerized or the networks replaced with computer networks.

Network structure

In general, every telecommunications network conceptually consists of three parts, or planes (so called because they can be thought of as being, and often are, separate overlay networks):

The control plane carries control information (also known as signalling).

The data plane or user plane or bearer plane carries the network's users traffic.

The management plane carries the operations and administration traffic required for network management.

Example: the TCP/IP data network

The data network is used extensively throughout the world to connect individuals and organizations. Data networks can

be connected to allow users seamless access to resources that are hosted outside of the particular provider they are connected to. The Internet is the best example of many data networks from different organizations all operating under a single address space.

Terminals attached to TCP/IP networks are addressed using IP addresses. There are different types of IP address, but the most common is IP Version 4. Each unique address consists of 4 integers between 0 and 255, usually separated by dots when written down, e.g. 82.131.34.56.

TCP/IP are the fundamental protocols that provide the control and routing of messages across the data network. There are many different network structures that TCP/IP can be used across to efficiently route messages, for example:

- wide area networks (WAN)
- metropolitan area networks (MAN)
- local area networks (LAN)
- campus area networks (CAN)
- virtual private networks (VPN)

There are three features that differentiate MANs from LANs or WANs:

The area of the network size is between LANs and WANs. The MAN will have a physical area between 5 and 50 km in diameter.^[2]

MANs do not generally belong to a single organization. The equipment that interconnects the network, the links, and the MAN itself are often owned by an association or a network provider that provides or leases the service to others.^[2]

A MAN is a means for sharing resources at high speeds within the network. It often provides connections to WAN networks for access to resources outside the scope of the MAN.^[2]

Security Management in Telecom Networks

The import of telecom equipment from other countries that are antagonistic to a state's strategic interests may lead to supply chain contamination by means of embedded logic bombs and malware. The dependence on telecommunication networks and the critical role that they play in the economic growth of a country has led to government regulations in the telecom industry, which include requirements for ensuring the security of the telecom equipment and networks.

The interconnection of the PSTN networks of fixed and mobile phone systems and the next generation network has increased the attack surface of the telecom networks. The

wide range of end-user devices that can now connect to the telecom networks has added to the complexity of the networks, thereby increasing the risks and vulnerabilities as well.

The security threats and challenges to the telecom network listed in the Section 3 are indicative of the risks to these networks. As noted, the consequences of not implementing adequate security measures to deal with these could be heavy.

Several international standard development organisations like ITU, ISO/IEC, 3GPP, 3GPP2 and ETSI have prescribed standards that are applicable to telecom networks. Some of the most prominent standards that include requirements/guidelines for the security of telecom networks are listed in Table 2. Also, many countries have legislations and regulations that the telecom operators must comply with, which may require the adoption of specific security standards.

Telecom operators should adopt a robust, managed security programme to ensure that their networks are protected against malicious attacks, both external and internal, while also ensuring compliance to the local regulatory environment. This requires a holistic approach to implementing security measures, based on globally accepted security standards and best practices. [5] Provides an overview of security requirements, threat identification frameworks and guidelines for risk mitigation Incident organisation and security incident handling

Telecom Network Security Management

A multi-pronged approach to security should be adopted by telecom operators to address the current and future security challenges. Industry-recognised standards, best practices and technologies must be adopted to build a robust security programme. In addition, all applicable legal and regulatory requirements should also be considered.

Adopting a Security Framework

Organisations develop and implement security policies and procedures to address the security requirements for their environment. However, to be effective, these policies and procedures should be tightly coupled, and supported by industry-accepted guidelines, standards and best practices. There also should be a risk-based approach while developing these policies to ensure that the security measures are adequate to the address the perceived business risks.

Several IT Frameworks available today, like COSO, COBIT, ITIL, ISO27001 and others, can be adopted to

formulate a security programme. The ISO 27001:2005 standard is one of the most widely accepted security standards across industries. This provides a model for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an Information Security Management System (ISMS). For the telecom industry, this is further supported by ISO 27011:2008, which provides guidelines on information security management for telecommunication networks (jointly developed along with ITU-T).

The ISO 27001 standard is based on the Plan-Do-Check-Act (PDCA) model, which is applied to all ISMS processes, as illustrated in Figure 1. This PDCA model ensures that there is a continued focus on the security programme, and that it is not a one-time activity.

Future Scope

Networks are so commonplace within organizations, network security is important for all administrators. Maintaining good network security is a full-time task that has to involve the cooperation of all employees within an organization.

Because network security is so important, and involves all aspects of day-to-day operations, it is important that security policies be communicated from the top down, and that all managers are involved in the planning of network security policies.

Many organizations, especially small ones, don't feel that they need to worry about network security. The truth is, any organization that is publicly connected to the Internet has to make an effort to secure its border.

2. CONCLUSION

As telecom technologies advance are more widely deployed, it is essential that telecom operators put their best foot forward to secure their networks and services. It is also critical that they conduct periodic risk assessments of their networks and tweak their security programmes to adapt to the ever-changing security environment. As new vulnerabilities are discovered, new threats emerge, and security products evolve, operators need to take judicious decisions to choose the right security solutions and methodologies, in line with their risk appetite.

REFERENCES

- [1] Convergence and Next Generation Networks, Ministerial Background Report (OECD), 2007
- [2] “Security in the Traditional Telecommunications Networks and in the Internet”, Markus Isomäki, November 1999
- [3] NIST Special Publication 800-13, Telecommunications Security Guidelines for TMN, October 1995
- [4] “A Guide to 3rd Generation Security”, 3GPP TR 33.900 version 1.4.0
- [5] Security in Telecommunications and Information Technology, An overview of issues and the deployment of existing ITU-T recommendations for secure telecommunications, ITUT, June 2006
- [6] Unknown Vulnerability Management for Telecommunications, Anna-Maija Juuso and Ari Takanen, Codenomicon, February 2011

An Artificial Neural Network Based Approach of Rejecting Static Clutter from the Covariance Matrix Containing Low Velocity, Small RCS Target in Case of Monostatic Airborne Radar

M Chakraborty¹, P Karmakar², B. Maji³, D. Kandar⁴

¹Surendra Institute of Engineering & Management, Siliguri, India

²National Institute of Technology, Durgapur India

³SKP Engineering College, Tamilnadu, India

Abstract: Clutter, static or dynamic exhibits a major problem for outdoor Radar operations and there is no such standard measure devised yet to reject it online. Moving Target Indicator (MTI) is a proven technique which currently used in airborne Radar, the advantage of the Doppler frequency in MTI is used in ITS and the desired results has been achieved. This paper presents a novel approach of strong & effective rejection of static ground clutter from slowly moving object implemented with the help of ANN technology for GMTI airborne platform.

Index Terms - ANN, CPI, GMTI, KA- STAP

1. INTRODUCTION

In modern war field scenario the most popularized remote sensing application is based on networked ground moving target indicator (GMTI) Radar [1]. The Radars within the same network have the concurrent sensing & communication technology [2,3] with each other. The GMTI Radar is operational with the scheme that it would sense firstly the ground slow moving object (i.e., war tanker, soldiers and others) and necessarily it can discriminate the static clutter from the slow moving object[4] with the help of Artificial neural network[5] based RADAR receiver. For implementing ITS the vehicles will have to be smart enough to detect friends & fro of each other. Vehicles will sense each other (both slow moving & fast moving) & pass this sensing information to each other instantaneously. After the remedy of the static clutter, the velocity information along with the positional data of the slow moving object are fused and then promptly communicated to another high velocity GMTI Radar or to a ground fixed navigation point. As in another possibility, any two GMTI Radars move with very high velocity, the remotely sensed information is communicated to a third GMTI Radar so that it can make its trajectory precise to meet the objective of the ongoing attacks towards the enemy Radar. In considering these effects, one should intuitively expect that the received echo from a scatterer will fluctuate from pulse to pulse during a

coherent processing interval (CPI). This fluctuation will cause a broadening of the clutter ridge [9] and attempts to suppress this broadened ridge through the use of STAP [6] which will require additional degrees of freedom as the total clutter covariance matrix rank is increased. Space-Time Adaptive Processing (STAP) techniques promise to be the best means to detect weak targets in severe, dynamic, interference scenarios including clutter and jamming. STAP algorithms are only now being used for Ground Moving Target Indication (GMTI) from an air- or space-borne platform. Low velocity ground targets lie extremely close to the main beam clutter in Doppler, making detection difficult. The need to detect small, slow moving, ground targets in stressful interference environments drives research in adaptive processing techniques for GMTI. [7].

2. BACKGROUND

In an airborne early warning prototype for detecting targets buried in jamming and clutter interference multiple channels are input into the real-time signal processor to focus the energy in the direction of the target while simultaneously canceling the intentional jamming interference and the clutter interference generated from the motion of the aircraft. The properties exploited to make the target extraction from these high levels of interference are angle and Doppler. These properties have led to this technique's common reference as space-time adaptive processing (STAP)[8].

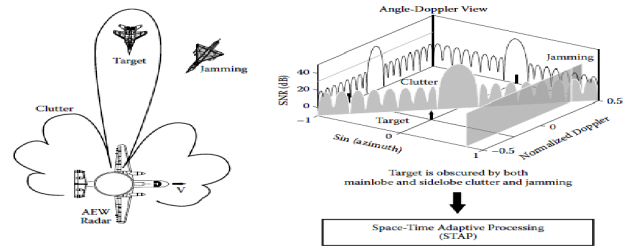


Fig. 1. Airborne Early Warning RADAR

The target in fig 1 is assumed to be present on the main beam of the array at zero-degree azimuth angle. Therefore, the competing jamming and clutter are assumed to be competing with the target at a given Doppler but arriving from a different angle[10] .

Problem defined

In designing a MVDR beam former receiver based on STAP [M. Chakraborty *et al*, “Designing a MIMO Digital MVDR Beam Former using STAP Processing for Adaptive Steering of Antenna Beam” International Journal of Mobile Communication & Networking. ISSN 2231-1203 Volume 2, Number 1 (2011), pp. 51-56], using MATLAB/Simulink. The existing problem in the above work done is to distinguish the ground static clutter from the slow moving object.

Proposed Solution

The difficulty in this regard, may be tackled by supervised ANN based approach where the desired network was trained with range & cross range profiles data within the operating constraint of the GMTI RADAR initially, then the network was simulated with different data of range & cross range profiles which generates the practical covariance matrix of the environment. After comparing the estimated RCS from simulated network with the generated covariance matrix, slow moving clutter can be rejected and radar target signature can be generated according to the fig 2.

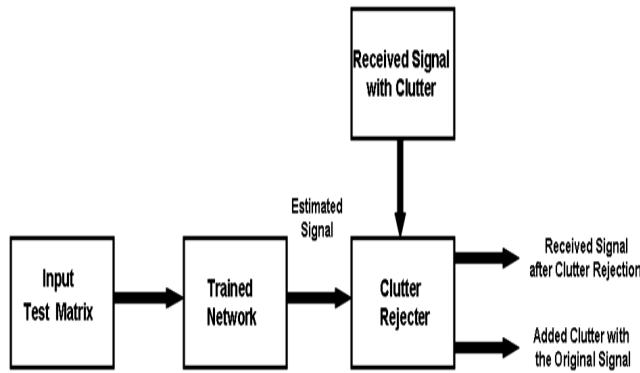


Fig. 2. Arrangement for Clutter Rejection

3. IMPLEMENTATION

The implementation was divided into different phases, such as:

1) Data collection, 2) Simulating the network, 3) pre-processing & post processing of data , 4) Clutter modeling, 5) Clutter Rejection. Figure 3 and figure 4 are showing the data used for network training and network training arrangement respectively.

Range & cross range profile data used for training		
	From	To
Range profile data [Frequency values]	22 GHz	29 GHz
Cross Range profile data [Aspect Angle]	0 degree	6 degree

Fig. 3. Table showing the data used for training the network.

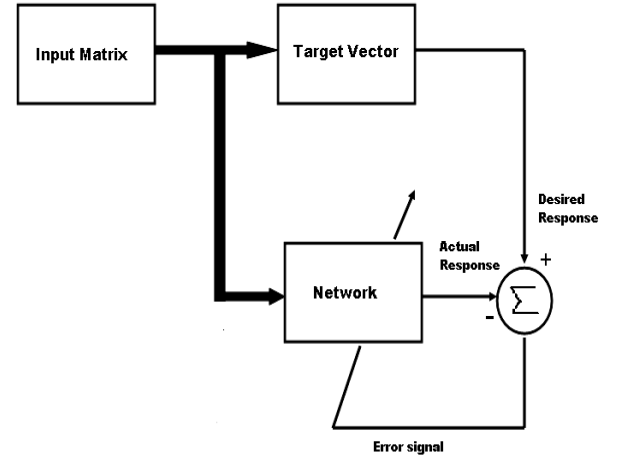


Fig. 4. Arrangement for Network training

The model for clutter used here is,

$$\sigma_{clutter} = \frac{\sin(\theta) \cos(\theta)}{\sec(\theta) \tan(\theta)} \times \frac{\sqrt{f}}{2\pi}$$

Where,

$\theta \rightarrow$ aspect angle

$f \rightarrow$ frequency

Clutter RCS are taken in range and cross-range profile for the same values of frequency and aspect angle. To reject Clutter first RCS values for slow moving [20 meter/ sec] target were stored in range and cross-range profile for the frequency and aspect angle values named as ‘Original Signal’. Then practical covariance matrix i.e., ‘Original Signal’ with added ‘Clutter’ signal is received at the Radar receiver.

$$Received\ Signal = Original\ Signal + Clutter\ Signal.$$

Now, the trained Network was simulated for the same values of frequency and aspect angle and the signal then generated is ‘Estimated Signal’. ‘Received Signal’ and ‘Estimated Signal’ when compared, the low velocity clutter is rejected and returns ‘Original Signal’. [Added Clutter = Received Signal – Estimated Signal.] Signal obtained by rejecting this

‘Added Clutter’ from ‘Received Signal’ is the ‘Clutter rejected Received Signal’ or target signature.

4. RESULTS

Table 1 is showing the initialization of different network parameters before training

Table 1: Initial training parameters and assigned corresponding data

Learning rate	0.1
Epochs	Infinity
Error-goal	10^{-6}
Validation check	Infinity
Time	Infinity
Hidden Layer activation function	Tan sigmoid
Output Layer activation function	Linear

Tables 2 and 3 are showing values of different network parameters during and after training,

Table 2. Error gradient reached during different phases of network training

During training	0.2×10^{-6}
During validation	2×10^{-6}
During testing	10^{-7}
Overall	7.59×10^{-6}

Table 3. Correlation coefficient ‘R’ for Regression analysis reached during different phases of network training

During training	0.99942
During validation	0.99935
During testing	0.9996
Overall	0.99927

Performance Plot

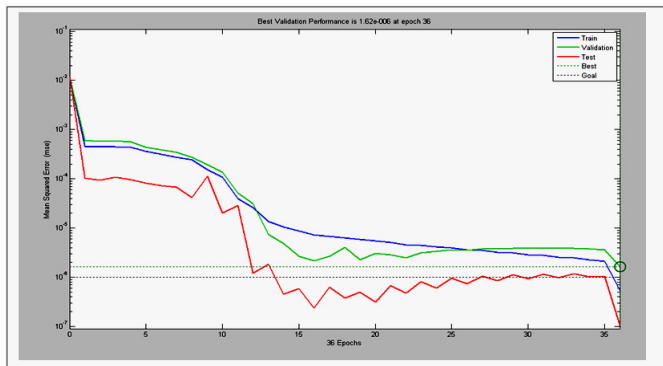
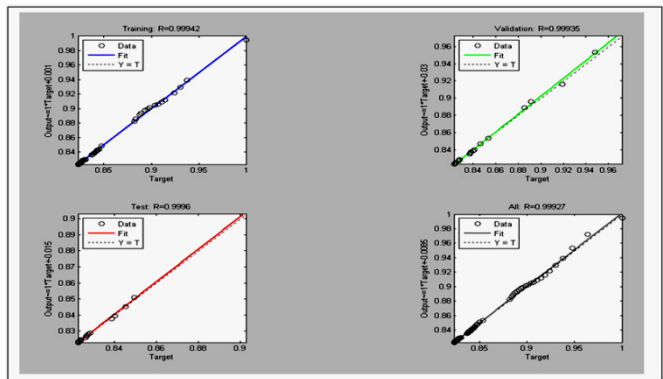


Fig. 5. Performance Plot

Regression Plot



From the Performance plot it is clearly seen that error-goal (blue colored line) is reached during training at 36 epochs. Regression plot is to perform a linear regression between the network outputs and the corresponding regression targets, from the Regression plot it is clearly seen that the output tracks the targets very well for training, testing, and validation, and the R-value is over 0.99927 for the total response.

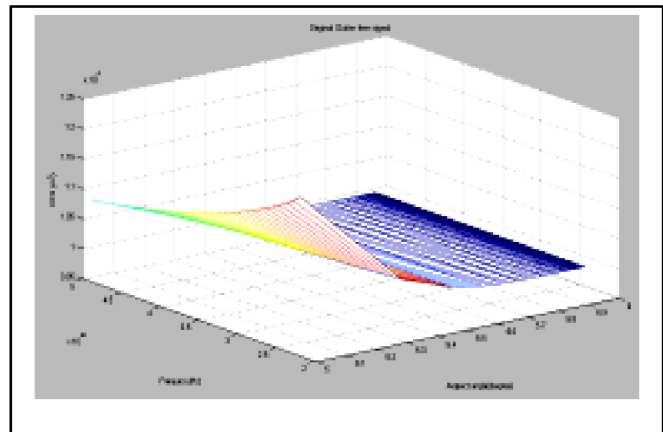


Fig. 6. original clutter free signal

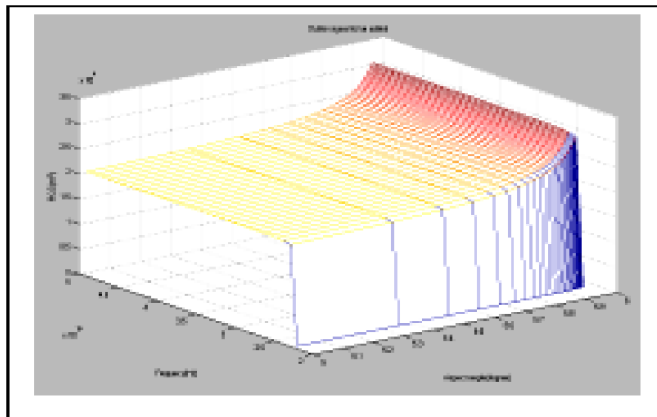


Fig 7: Clutter signal to be added

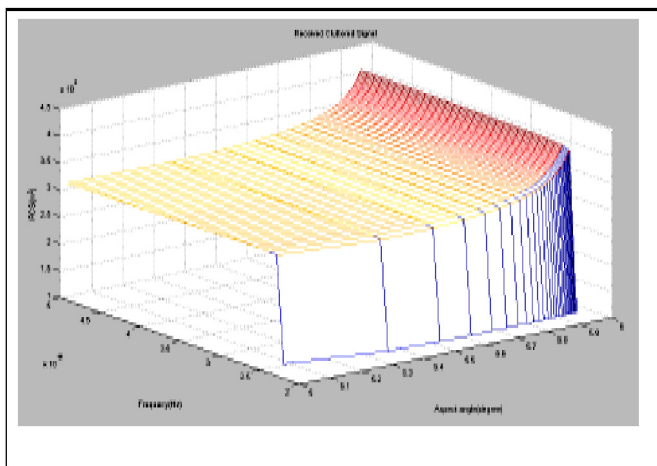


Fig 8: Received signal with Clutter

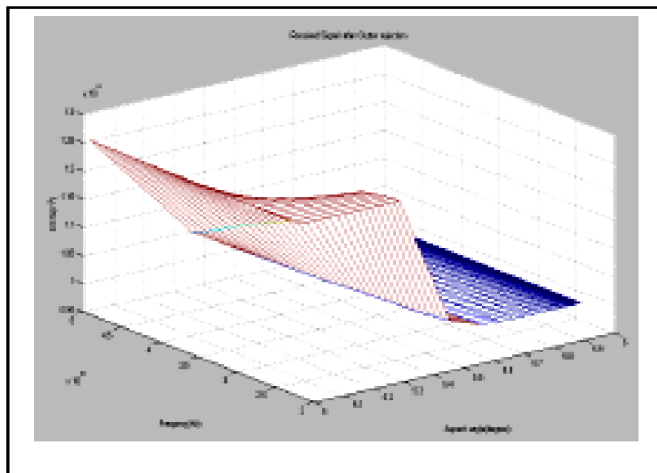


Fig 9: Received Signal made clutter free with help of ANN.

5. ANALYSIS & CONCLUSION

This technique of Rejection of Clutter comes under the function approximation task of ANN. Herein the problem we have tried to implement the artificial Neural Network Technology in accordance with KA-STAP for differentiating the slow moving target from fixed clutter. The objective of this review is to extend the advantages of STAP in the high velocity region to lower velocity targets by using KB/KA-STAP with ANN.

6. FUTURE WORKS

The works that remain unaddressed here is the hardware implementation of this ANN based STAP processor to detect human movement inside a house.

REFERENCES

- [1] Michael C. Wicks, "Knowledge-Based Control For Space Time Adaptive Processing", Air Force Research Laboratory, 2006, USA.
- [2] G. Lellouchl, H. Nikookar, "On the Capability of a Radar Network to Support Communications", Delft University, Netherlands.
- [3] M. Jamil, H. J. Zepernick, "On Integrated Radar and Communication Systems Using Oppermann Sequences", 978-1-4244-2677-2008 IEEE.
- [4] M. LESTURGIE, "Use of STAP techniques to enhance the detection of slow targets in ship borne HFSWR", A Report, dated- 14 APR 2005
- [5] A. Katidiotis, K. Tsagkaris, "Performance Evaluation of Artificial Neural Network - based Learning Schemes for Cognitive Radio Systems", University of Piraeus, Greece.
- [6] Wolfram Bürger, "Space-Time Adaptive Processing: Algorithms", RTO-EN-SET-086, <http://www.rto.nato.int/abstracts.asp>.
- [7] R. Adve, T. Hale, and M. Wicks, "knowledge based adaptive processing for ground moving target indication", http://www.comm.toronto.edu/~rsadve/Publications/KB_STA_P_DSP.pdf, dated- 23/01/13
- [8] Ke Sun, Hao Zhang, Gang Li, Huadong Meng, Xiqin Wang, "Airborne Radar STAP using Sparse Recovery of Clutter Spectrum", [www. http://arxiv.org/ftp/arxiv/papers/1008/1008.4185.pdf](http://arxiv.org/ftp/arxiv/papers/1008/1008.4185.pdf), dated – 23/01/2013
- [9] Z. Ning Zhou, "Space-time adaptive processing with multi-stage Wiener filter and principal component signal dependent algorithms", a thesis of MS in EE dept. California Polytechnic State University, dated- 23/01/13.
- [10] D. R. Martinez, R. A. Bond, M. M. vai, "High Performance Embedded Computing Handbook", Chapter 2.
- [11] Marc LESTURGIE, "Use of STAP techniques to enhance the detection of slow targets in shipborne HFSWR", a report, 14 APR 2005, ONERA.

Plug and Play VOIP with Raspberry Pi for Small Scale Industries

Chirag Gohel¹, Himanshu Madhavani²

¹Assistance Professor, ²Student

^{1,2}Marwadi Education Foundation, Rajkot

¹chiragkgohel@gmail.com, ²madhavanihimanshu@yahoo.co.in

Abstract: In today's world IT Industries are under Pressure to deliver to get more and more Functionally and speed at low cost , But as the number of devices increases in the Industries which requires more workspace and cost.

The Software that are used by Open Source Hardware also follows terms of the free and open source software (FOSS). Extensive Efforts has been taken in order to run Hardware by using Open Source Software's. So one of the newly came Open Source Hardware named "Raspberry pi "which runs with an Linux Environment with 512MB of Ram , ARM v 6.

Asterisk is one of the world's leading open source telephony software/system, can be easily Deployed on Internet as well as Local network. Asterisk offers the same quality of services as propriety software's. An Asterisk system can be considered as replacement of traditional PBX systems in the LAN environment.

Asterisk enables Voice-Video-Data communication from personal phone to SOHO to Enterprise Communication System. Good Quality of Service (QoS) for Voice-Video can be achieved using Asterisk Open Source Telephony System.

The paper Focuses on Deployment of VOIP server on small Embedded Device-Raspberry pi which can be used as plug and play Portable Device, Which can be used to overcome communication problem in case of Disaster. One can also make it low cost, small office VoIP Server for internal communication.

Keywords: Raspberry Pi, Asterisk, IP Telephone, Linux Server, Embedded Communication, Portable VoIP.

1. INTRODUCTION

Raspberry Pi is one of new Open Source hardware which uses ARMv6 Processor with 512MB of RAM with Linux as working Environment.

Asterisk, Open Source Telephony Service using platform as Linux. It is used for good quality of Telephony Services over IP based Network. Asterisk offers different sharing and voice over Ip over the internet as well as local LAN. Asterisk is developed by following Open Source Software

Licenses, So can be modified easily according to use as well as for Low Budget Communication System.

Raspberry Pi and Asterisk turns a small Linux Telephony server which can work for easy and inexpensive System. Raspberry Pi will serve many VOIP devices by adopting Asterisk as a Service. Asterisk when combined with Raspberry pi creates a PBX at a fraction of the price of traditional PBX systems. Asterisk can be deployed in fraction and also it provides best quality service in parallel to legacy systems at low cost. By using Asterisk Service one can easily connect SOHO VoIP with the rest of the world, where people are using Asterisk to fulfill their communication need. [2]

By combining Asterisk with Raspberry Pi – Open Source Hardware, it is very easy to deploy VoIP server for SOHO. As Raspberry Pi is one of the Plug and Play Embedded Device, it also provides the portability for the deployment of VoIP Service. We can take the advantage of Quick uptime to establish System for any emergency or disaster recovery system.

2. REVIEW OF LITERATURE

Asterisk-An Open Source Software/Service deployment with aspect of small-scale implementation discussed in [3]. Raspberry Pi – An Open Source Hardware, Embedded Mini Computer with Open Source Platform allows to deploy VoIP and test the environments. Also present some important theoretical and experimental result regarding setting up a Raspberry Pi for basic functioning. [10]. Tested with the deployment of VoIP (voice over Internet protocol) server with the well known open source software - Asterisk on Raspberry Pi, this research also enlightening implementation of DHCP and Asterisk server setup configuration as well. VoIP deployment using Raspberry Pi [12], in this research deployment aspects of Asterisk over Raspberry Pi for Voice exchange, developing a prototype for Plug and Play VoIP server for SOHO, to ready communication service during disaster in a quick time, to setup Communication System for temporary base. Also connected and configured the client

end with the soft phones and tested successfully the communication using VoIP with Raspberry Pi.

In [4] connection of the clients to the server with the help of IAX protocols also discussed and deployment configuration discussed. Important features of Asterisk deployed in [5], the services generally associated with an Asterisk based Voice Exchange i.e. conferencing, paging and voice mailing.

Communication System with Raspberry Pi

SIP and H.323 makes Asterisk as one of the major telecommunication systems in both US and European countries. Asterisk can easily connect with the existing IP based network for the integration of voice & data. It also support traditional phone network with the conversion from analog to digital. Client end can also be supported by regular telephone as well Soft/Hard IP Phone too.

ARM 1176JZFS 700 MHz processor with Broadcom BCM2835 System on chip (SoC) makes Raspberry Pi a small, hand held minicomputer. Using overclocking facility user can easily reach up to 1GHz, without any damage to the hardware. Up to 512 MB RAM shared with GPU makes Raspberry Pi a great device to compete with the existing hardware in the market. Debian on ARM platform as an startup Operating System gives a different look to this device.

Asterisk with Raspberry Pi creates and different plug and play communication infrastructure, able to exchange Voice-Video-Data over Intranet as well as Internet. Calls over Asterisk seem to be much cheaper in cost, better in Quality and much more functionality over the traditional telecommunication system in a shortest uptime. Portability gives one more dimension to the Raspberry Pi for moving the whole infrastructure from one place to another in short time duration.

Asterisk allows centralized dictionary for contacts, Voice-Video conferencing, Call Parking, Voice Mail, Web based access to voice mail and many other facilities. Asterisk's flexible dial plan allows seamless integration of IVR and PBX functionality. Asterisks Features can be implemented with different logical dial plans. TCP, UDP RTP, RSVP are major protocols which works on IP to make Asterisk service success in the real time communication system.

Raspberry Pi with Networking Support in the Operating System allows user to provide all the networking related services. Using DHCP Service in Raspberry Pi for providing IP Addresses to the client at the time of request for VoIP Service makes infrastructure more smooth in terms of plug and play facility.

Inter-Asterisk Exchange (IAX) is a Voice over IP protocol specific for Asterisk to exchange the information between two Asterisk Servers and allow to communicate easily. IAX allows Asterisk to merge voice and data traffic seamlessly across disparate networks. When using Packet Voice, data like URL information and images can be sent in-line with voice traffic. This supports advanced integration of voice and data that is not available in legacy systems. Asterisk provides a central switching core, with four APIs for modular loading of telephony applications, hardware interfaces, file format handling, and codec. Asterisk provides transparent switching between all supported interfaces. With this Asterisk ties together diverse telephony systems into single switching network.

Service List for Raspberry Pi with Asterisk

In our testing bed environment we tried and tested following services on Raspberry Pi with Asterisk Service. Call Waiting, Call Transfer, Call conferencing etc provides more features on the client end device. Music on Hold, Call Parking, Roaming extensions, SMS Servicing and Voice Mail enables enterprise level features for the IP based Telephony System on Open Source Platform.

- Call Detail Records
- Call Forward
- Call Monitoring
- Call Transfer
- Call Waiting
- Caller ID
- Caller ID Blocking
- Database Store / Retrieve
- Privacy
- Remote Call Pickup
- Roaming Extensions
- Talk Detection
- VoIP Gateways
 - i. Voicemail
 - ii. Visual Indicator for Message Waiting
 - iii. Voicemail Groups

Testing Environment: Asterisk & DHCP on Raspberry Pi Small Scale Deployment Raspberry Pi

As shown in the illustration of this paper, Asterisk Server on Raspberry Pi can easily connect with Soft and Hard Phones using IP based network for providing easy Voice

Communication [4]. IP phone can be easily connected with the Raspberry Pi. Any Standard IP Soft/Hard Phone, running SIP or H.323 can easily communicate with Raspberry Pi to use Asterisk Service. Any call can be established between Soft and hard phone, no matter where that phone is in IP based Network. It is assured that connections are secure and that unauthorized users are excluded. Any phone controlled by Asterisk system can call any other VoIP.

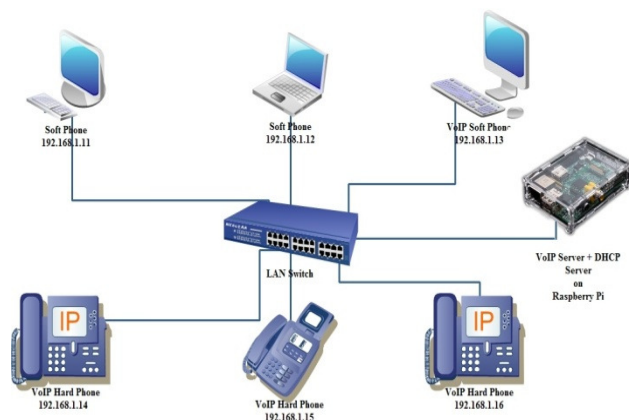


Fig. 1. Small Scale Deployment of Raspberry Pi

Interconnection of Raspberry Pi

As shown in the following illustration, Different Raspberry Pi can be connected with each other and all users can easily place a call in the full communication network. Once Raspberry Pi can connected users should be able to place calls between two or more than two corporate office placed at different geographical area. It is one kind of extension to the Simple Asterisk deployment, where one user can talk with the other who is in the other corporate office without any cost. Just to make connection between different Raspberry Pi with Asterisk as a Service at a single point will allow calls to route from one office to another.



Fig. 2. Interconnection of Raspberry Pi.

Asterisk and Other Proprietary Telephone System

For small to large organization, there are various telephony systems available in the market. All the proprietary systems build with its own hardware and software. To have monopoly in the market for business, proprietary hardware

and software are not compatible to work with other company hardware and infrastructure.

Manufacturers usually sell the largest systems themselves, through a dedicated sales force. A dedicated sales force is, of course, expensive. The cost of this sales force and all the support behind the sales force is included in the price you pay for your telephone system. Anything smaller than the very largest systems is usually sold through representatives or distributors. The smallest systems are typically available through representatives or distributors

Call Pricing

The price paid in the Proprietary telephone systems includes cost of manufacturing and distribution. The price of such Telephony System may be are high as it include profit of distribution chain, distributor, retailer so their prices are very high. So for the solution to this Asterisk is designed and prepared with commodity hardware.

As Asterisk as well as Raspberry pi Systems are inexpensive. To use this System as service user needs interface board or normal soft phones that support Telephony and can be hooked up to an incoming and outgoing calls.

While in the propriety systems as you add features it will cost you more /feature. (E.g. conferencing, call waiting). Each Propriety Systems will be designed for certain number of users, adding or modifying the system will again cost more money. A small Raspberry pi that will turn to an Asterisk Server that can support Surprising features and number of users [7].

Easy Implementation and Acquiring Innovative Services

A small Raspberry pi device lots of features beyond its size. Asterisk when deployed on Raspberry pi will convert it to small plug and play device which can provide inexpensive telephony system at runtime. Asterisk contains many features which are found with high-end propriety systems. A community of users has grown up around Asterisk [8]. Asterisk is developed in such a way that we can easily add the feature we want at runtime, since it is Open Source. As in the same way Raspberry Pi can be added with the features like wired as well wireless services by adding different modules to it [11]. Asterisk and Raspberry community helps every user and helps us to add features to it.

This system provides better flexibility, durability, easy customization. For example, Raspberry and Asterisk has scripting system, this will help us to add any features or customization as per needs. For example, if you want that the system should make automated call at particular time then we can write a script for it.

Geographically Independency and Quality of Service (QoS)

This System does not have any relation with the network or the place where it is used [9]. User can have this system any where irrespective to location. For example, in case of earthquake system can be deployed in couple of minutes and telephonic conversations can be created in network easily. As well this system at one place can be connected to similar system which can communicate with each other. For example, Office at A location can easily get connected to another office at location B via Internet line. Thus in the same way this system as a part of Open Source helps in reducing the costs as well as provide better flexibility in case of disasters for telephonic conversations.

SIP Scenario for System

Steps for create SIP Registration Name in Asterisk
SIP/device name where device name is defined in a section below.

SIP/username@domain to call any SIP user on the Internet.

context=default	Default context for incoming calls
allowguest=no	Allow or reject guest calls (default is yes)
allowoverlap=no	Disable overlap dialing support. (Default is yes)
allowtransfer=no	Disable all transfers (unless enabled in peers or users) Default is enabled
realm=mydomain.tld	Realm for digest authentication defaults to "asterisk". If you set a system name in asterisk.conf, it defaults to that system name Realms MUST be globally unique according to RFC 3261 Set this to your host name or domain name
bindport=5060	UDP Port to bind to (SIP standard port is 5060) bindport is the local UDP port that Asterisk will listen on
bindaddr=0.0.0.0	IP address to bind to (0.0.0.0 binds to all)
srvlookup=yes	Enable DNS SRV lookups on outbound calls Note: Asterisk only uses the first host in SRV record. Disabling DNS SRV lookups disables the ability to place SIP calls based on domain names to some other SIP users on the Internet
domain=mydomain.tld	Set default domain for this host If configured, Asterisk will only allow INVITE and REFER to non-local domains Use "sip show

	domains" to list local domains
pedantic=yes	Enable checking of tags in headers international character conversions in URIs and multiline formatted headers for strict SIP compatibility (defaults to "no")
maxexpiry=3600	Maximum allowed time of incoming registrations and subscriptions (seconds)
minexpiry=60	Minimum length of registrations/subscriptions (default 60)
defaultexpiry=120	Default length of incoming/outgoing registration
t1min=100	Minimum roundtrip time for messages to monitored hosts. Defaults to 100 ms
notifymime-type=text/plain	Allow overriding of mime type in MWI NOTIFY
checkmwi=10	Default time between mailbox checks for peers
buggy-mwi=no	Cisco SIP firmware doesn't support the MWI RFC fully. Enable this option to not get error messages when sending MWI to phones with this bug.
vmexten=voicemail	dialplan extension to reach mailbox sets the Message-Account in the MWI notify message defaults to "asterisk"
disallow=all	First disallow all codec's
allow=ulaw	Allow codecs in order of preference
allow=ilbc	see doc/rtp-packetization for framing options
Video support=yes	Turn on support for SIP video. You need to turn this on in the this section to get any video support at all. You can turn it off on a per peer basis if the general video support is enabled, but you can't enable it for one peer only without enabling in the general section.
Maxcall bitrate=384	Maximum bitrate for video calls (default 384 kb/s) Videosupport and maxcallbitrate is settable for peers and users as well

----- RTP timers -----

These timers are currently used for both audio and video streams. The RTP timeouts are only applied to the audio channel.

The settings are settable in the global section as well as per device

rtptimeout=60	Terminate call if 60 seconds of no RTP or RTCP activity on the audio channel when we're not on hold. This is to be able to hang-up a call in the case of a phone disappearing from the net, like a power loss or grandma tripping over a cable.
rtpholdtimeout=300	Terminate call if 300 seconds of no RTP or RTCP activity on the audio channel when we're on hold (must be > rtp timeout)

```
Configuration of DHCP server ( dhcpd.conf )
subnet 192.168.1.1 netmask 255.255.255.0 {
option routers 192.168.1.1;
# Unknown clients get this pool.
pool {

max-lease-time 300;
range 192.168.1.2 192.168.1.19;
allow unknown-clients;

}
```

```
[himanshu]
username=himanshu
type=friend
host=dynamic
context=home
secret=123456
nat=no
allow=all
```

Create Extention No and Mapping of Name and No.

exten => 2001,1,Dial(SIP/himanshu)

If 2001 is the extension no for user. 1 will be the priority at the time of matching the extension no in the database.

Dial (SIP/HIMANSHU) provides the protocol match for the given extension and username.

```
Configuration at Asterisk Server [Site A]
[general]
bindport = 4569 ; Port to bind to (IAX is 4569)
bindaddr = 0.0.0.0 ; Address to bind to (all addresses on machine)
disallow=all
mailboxdetail=yes
[chirag]
```

```
type=friend
username=chirag
secret=123456
auth=plaintext
host=192.168.2.3
Context=fromiax
peer context=fromiax
qualify=yes
trunk=yes
```

Configuration at Asterisk Server [Site B]

```
[general]
bindport = 4569 ; Port to bind to (IAX is 4569)
bindaddr = 0.0.0.0 ; Address to bind to (all addresses on machine)
disallow=all
mailboxdetail=yes
[vinod]
type=friend
username=XYZ
secret=123456
auth=plaintext
host=192.168.1.2
context=fromiax
peer context=fromiax
qualify=yes
trunk=yes
```

Abbreviations Meaning (used for asterisk Configuration)

username	The username is the name that is stored in "iax.conf" from where we can find the peering information
secret	This is the connection password that is used for authentication.
host	Host is the IP address of the server bar. There will be no need of registration when we specify the IP addresses.
Dial plan	We would use a file that be used to add login and passwords , so we don't have to remember the plans.
Qualify	This option is used to generate couple of bits of ping between devices which will help us to know connectivity.
Context	It is context which would accept the extension bar by sending us when someone try to call us.
Peer context	This context is used to avoid typing context in the Dial() , it is useful to write strings easily.
Trunk	This function is useful when more then one call is active between two devices , they will sent in such a way that will save IP overhead and we will get better quality.

3. CONCLUSION

This is not the end of the Open source telephony system. By implementing this system we learn about many capabilities and functions of Raspberry pi and Asterisk, which can be used to implement fully IP oriented telephony system.

Asterisk enables Voice-Video-Data communication from personal phone to SOHO to Enterprise Communication System. Good Quality of Service (QoS)

In our current scenario we had used single server, our future expansion would be connecting other PSTN lines and other Raspberry pi Servers which will help us to make connectivity globally. Also we would connect with Inter Asterisk Exchange (IAX) which would help to connect SIP call between two devices at remote places.

Thus Raspberry Pi and Asterisk provides plug and play voice over IP using internet at low cost. Also it would be helpful at time of disaster when regular cell phones don't work, and provide good Quality of service.

REFERENCES

- [1] Mark Spencer, "Introduction to the Asterisk Open Source PBX", Presented first at Libre Software Meeting 2002, France
- [2] Paul Mahler, "VoIP Telephony with Asterisk", ISBN 09759992-0-6
- [3] Alam, M.Z. Bose, S. Rahman, M.M. Abdullah Al-Mumin, M., "Small Office PBX Using Voice over Internet Protocol (VOIP)" The 9th International Conference on Advanced Communication Technology, 2007 page: 1618 - 1622
- [4] Qadeer, M.A.; Imran, A, "Asterisk Voice Exchange: An Alternative to Conventional EPBX", International Conference on Computer and Electrical Engineering, 2008. ICCEE 2008. Page(s):652 – 656
- [5] Imran, A.; Qadeer, M.A, "Conferencing, Paging, Voice Mailing via Asterisk EPBX", International Conference on Computer Engineering and Technology, 2009. Page(s):186 – 190
- [6] Jim Van Meggelen, Leif Madsen, and Jared Smith, "Asterisk: The Future of Telephony", O'Reilly Media, 2007
- [7] Nir Simionovich, "AsteriskNOW", Packt Publishing, 2008
- [8] Montoro, P.; Casilari, E., "A Comparative Study of VoIP Standards with Asterisk Fourth International Conference on Digital Telecommunications, 2009. Page(s): 1 – 6
- [9] Konstantoulakis, G.; Sloman, M., "Call Management Policy Specification for the Asterisk Telephone Private Branch Exchange", Eighth IEEE International Workshop on Policies for Distributed Systems and Networks, 2007. Page(s): 251 – 255
- [10] Quick Start guide: The Raspberry pi – Single Board Computer product documentation, http://elinux.org/RPi_Hardware_Basic_Setup.
- [11] "Raspberri pi wifi adapter testing" from element14 raspberri pi forum pages <http://www.element14.com/community/docs/DOC-44703/1/raspberri-pi-wifi-adapter-testing> .
- [12] "Dhcp server configuration on debain system" <http://www.debianhelp.co.uk/dhcp.htm>

VoIP over Office Network

Sushil Kumar

*School of Information and Communication Technology
Gautam Buddha University, Greater Noida, Uttar Pradesh - 201 310, INDIA
sushilkumar0108@gmail.com*

Abstract: Many organisations are moving away from traditional separate voice and data networks to converged networks based on the Internet Protocol. Voice over IP (VoIP) technology provides the capability of transporting voice calls across the same Ethernet office infrastructure used for data and wireless technology gives workers the mobility everyone expects when using a telephone system both within the office and when on the road via public Wi-Fi hotspots.

A wide range of low cost VoIP solutions are available based on the IETF standard SIP protocol. The Brekeke PBX/SIP server is an example of a Windows software based system that can replace traditional internal PSTN exchanges and provide enhanced telephony services such as call conferencing. VoIP client software 'Soft phones' such as 3CX and Xlite can be used, with the aid of headsets, to enable Windows computers to process SIP VoIP calls and some Wi-Fi Smartphones have built in VoIP clients or apps to do the same. Dedicated hardware SIP phones are also available (or adaptors for analogue phones) that plug directly into LAN sockets or may have Wi-Fi capability.

There are, however, some issues related to call quality and security, particularly over Wi-Fi, to be solved before VoIP is truly accepted as a PSTN/PBX replacement for many offices.

Index Terms: VoIP, IETF Standard, SIP Protocol, PBX Server, Wi-Fi, PSTN.

1. INTRODUCTION

Voice over IP (VoIP) technology provides the capability of transporting voice calls across the same Ethernet office infrastructure used for data and wireless technology. It gives office people the mobility when using a telephone system both within the office and when on the road via public Wi-Fi hotspots. In VoIP first the Analog to Digital Converter converts analog voice to digital signals. Now the bits have to be compressed in a good format for transmission over internet.

Then the voice packets are inserted in data packets using a real time protocol (RTP over UDP over IP). SIP (Session Initiate Protocol) is needed for signaling between terminal units. At the Receiver site we have to disassemble packets, extract data, then convert them to analog voice signals and send them to sound card (or phone). All that must be done in a real time fashion because we cannot wait for too long for a

vocal answer. Voice over Internet Protocol (VoIP) is being increasingly deployed by enterprise IT departments as a cost effective replacement of PBX based telephone networks. The single biggest advantage of deploying VoIP is that the IT administrators have to maintain just one kind of network for both data voice applications resulting in significant cost savings. Moreover, the quality of VoIP calls has improved significantly over the past years as voice quality problems have been addressed in data networks as well as in VoIP devices. VoIP is also gaining popularity in the consumer space thanks to the availability of free PC to PC calling with softphones such as Skype [1]. Softphones allow a PC to be used as a VoIP phone and provide a lower cost alternative to mobile phones wherever broadband network connectivity is available.

Wireless LANs have been around for at least a decade. The popularity of WANs rose tremendously among notebook users after notebooks integrated WLAN solution became available along with Wi-Fi hot spots in popular places such as airports, hotels and coffee shops. Now a days, all kind of enterprises adopting WLANs as extensions of their corporate network beyond office spaces and cubicles [2].

As VoIP and WLAN deployments continue to gain the momentum in the enterprise areas, VoIP usage over WLAN is also expected to gain popularity, especially among notebook users with softphones. Supporting adequate quality for VoIP over WLAN has some unique challenges arising from the characteristics of WLANs. Typically, the bandwidth available over a LAN. Moreover, WLAN signal is susceptible to interference adjacent WLANs as well as other RF devices. This causes increased delays and sometimes losses for packets transmitted over WLAN. Another big challenge is that of Quality of Service (QoS) for VoIP traffic. WLAN products with QoS support have only been available in the last couple of years. This is because the base 802.11 specification [3] did not provide any mechanisms for separation or prioritization of traffic streams that have been different requirements for delay, packet loss rate, jitter etc. The 802.11e specification [4] from IEEE has been developed with the aim to provide Quality of Service (QoS) for a variety of applications such as voice, video and even prioritized data. Similarly, the Wireless Multimedia (WMM) specification [5] from Wi-Fi Alliance (WFA)

intends to bring out an early version of QoS support in 802.11 networks so as to allow the development of interoperable clients and Access Points (APs). Neither of these specifications however, provides specific recommendations for employing the QoS features for VoIP applications in an enterprise WLAN environment.

Voice over IP (VoIP) is susceptible to network conditions such as delay, jitter and packet loss. Delay is the time taken by a packet for travelling from one point to another (one way) in a network. Similarly, round-trip delay can also be measured. Users differ in their delay tolerance, but a good rule of thumb is to limit the one way delay to about 150 ms. VoIP packet delay comprises the following components:

- **Propagation delay:** Propagation delay is proportional to the speed of light and depends upon the physical distance between the two communicators.
- **Transport delay:** Transport delay occurs because of network devices such as routers, firewalls, traffic shapers, etc. The delay can either be constant or vary with the traffic.
- **Packetization delay:** This is a function of the CODEC speeds. Low speed CODECs such as G.723, take around 67.5 millisecond to convert analog signals into digital packets. An extra time is required because these CODECs have to compress the packets to reduce their size. High speed CODECs such as G.711 can packetize in approximately one millisecond.
- **Jitter buffer delay:** A jitter buffer helps to minimize the variations in the arrival times of the voice datagrams. However, sometimes in the event of excessive delay, packets have to be discarded.

Also, with more classification of packetisation delay, when coders/decoders (CODECs) in VoIP terminals compares voice signals, they introduce three types of delay:

- **Processing, or algorithmic, delay:** the time required for the CODEC to encode a single voice frame.
- **Look ahead delay:** the time required for a CODEC to examine part of the next frame while encoding the current frame.
- **Frame delay:** the time required for the sending system to transmit one frame.

Jitter is the variation in delay over time. If the delay of transmissions varies too widely during a VoIP call, the call quality is greatly degraded. The amount of jitter buffer tolerable on the network equipment in the voice path. The more jitter buffer is available, the more the network can reduce the effects of jitter. Packet loss is losing packets along the data path which severely degrades the voice application.

On deploying VoIP applications, it is important to assess all these three characteristics (delay, jitter and packet loss) for expected situations in the network in order to determine which of the available voice CODECS would perform well. The delay, jitter, and packet loss measurements can then aid in the correct configuration of prioritization, as well as setting other QoS mechanisms.

Standard bandwidth monitoring techniques and traffic analyzers can provide basic information on network traffic and help identify potential problems. The end systems can select different VoIP CODECS and simulate different call loads during real time tests. Different CODECS as G.711 and G.729, provide different sampling rates that affect packet size.

A. Speech CODECS

A CODEC is a software drivers or programs capable of encoding and decoding a digital data stream or signal. The word CODEC stands for "Coder-DECoder". A CODEC (the program) should not be confused with a coding on compression format or standard a format is a document (the standard), a way of storing data, while a codec is a program (an implementation) which can read or write such files. A CODEC encodes a data stream or signal for transmission, storage or decodes it for playback or editing. CODECs are used in video conferencing, streaming media and video editing applications. The analog to digital converter (ADC) converts its analog signals into digital signals, which are then passed through for digital transmission or storage. A receiving device then runs the signal through a decompressor, then a digital to analog converter (DAC) for analog display.

CODECs are software drivers or programs that can be used to encode the signal or data stream in a compact enough form that they can be sent in real time across the Internet using the bandwidth available. The conversion of an analogue waveform to a digital form is carried out by a CODEC. The CODEC samples the waveform at intervals and generates a value for each sample. These samples are typically taken 8000 times a second. The CODEC samples the waveform at regular intervals and generates a value for each sample. These samples are accumulated for a fixed period of time to create a frame of data. A sample period of 20 ms is common. Some CODECS use longer sample periods (such as 30 ms employed by G.723.1) or shorter ones (such as 10 ms employed by G.729a). The important characteristics of the CODEC are

- (i) The number of bits produced per second.
- (ii) The sample period defines how often the samples are transmitted.

These two characteristics define the size of the frame. For example, consider G.711 CODEC sampling at 20 ms. It generates 50 frames of data per second. G.711 transmits 64000 bits per second. So, each frame will contain $64000/50=1280$ bits or 160 octets.

CODEC algorithm seek to minimize the bit rate in the digital representation of a signal without an objectionable loss of signal quality in the process. High quality is attained at low bit rates by exploring signal redundancy as well as the knowledge that certain types of coding distortion are imperceptible because they are masked by the signal.

Models of signal redundancy and distortion masking are becoming increasingly more sophisticated, leading to continuing improvements in the quality of low bit rate signals. This section summarizes current capabilities in speech coding and describes how the field has evolved to reach these capabilities. It also mentions new classes of applications that demand significant improvements in speech compression and comments on how we hope to achieve such results.

B. PCM CODECS

Pulse Code Modulation (PCM) CODECS are the simplest form of waveform CODECS. Narrowband speech is typically sampled 8000 times per second and then each speech sample must be quantized. If linear quantization is used then about 12 bits per sample are needed, giving the bit rate of about 96 Kbps. This rate can be easily reduced by using non-linear quantization. This gives a bit rate of 64 Kbps and two such non-linear PCM CODECS were standardized in the 1960s. In America μ -law coding is the standard, while in Europe the slightly different A-law compression is used. Because of their simplicity, excellent quality and low delay both these codecs are still widely used today. G.711 is the mandatory minimum standard for all ISDN terminal equipment.

C. ADPCM CODECS

Adaptive Differential Pulse Code Modulation (ADPCM) codecs are waveform codecs which, instead of quantizing the speech signal directly like PCM codecs, quantize the difference between the speech signal and a prediction that has been made of the speech signal. If the prediction is accurate then the difference between the real and predicted speech samples will have a lower variance than the real speech samples, and will be accurately quantized with fewer bits than would be needed to quantize the original speech samples. At the decoder, the quantized difference signal is added to the predicted signal to give the reconstructed speech signal. The performance of the codec is aided by using adaptive prediction and quantization, so that the predictor and

difference quantizer adapt to the changing characteristics of the speech being coded. In the mid-1980s, the CCITT standardized a 32 Kbps ADPCM, known as G.721, which gave reconstructed speech almost as good as the 64 Kbps PCM codecs. Later in recommendations of G.726 and G.727, codecs operating at 40, 32, 24 and 16 Kbps were standardized.

D. iLBC

iLBC (internet Low Bitrate Codec) is a FREE speech codec suitable for robust voice communication over IP. The codec is designed for narrow band speech and results in a payload bit rate of 13.33 Kbps with an encoding frame length of 30 ms and 15.20 Kbps with an encoding length of 20 ms. The iLBC codec enables graceful speech quality degradation in the case of lost frames, which occurs in connection with lost or delayed IP packets.

2. VOIP SOFTWARE

There are a number of client software modules that allow users to transfer the voice data over the network using a selected vocoder.

A. Linphone

Linphone is a good graphical SIP softphone, with support for several codecs. It runs under Linux and works with the Gnome Desktop under Linux. It can also be used under KDE. Since version 0.9.0, linphone can be compiled and used without gnome, in console mode, by using the program called "linphonec". It works as simply as a cellular phone. Linphone includes a large variety of codecs (G711-ulaw, G711-alaw, LPC10-15, GSM, SPEEX and iLBC). Thanks to the Speex codec, it is able to provide high quality of voice even over slow Internet connections (such as 28k modems). It understands the SIP protocol, a standardized protocol from the IETF which is the organization that made most of the protocols used in the Internet. This guaranties compatibility with most SIP-compatible Web phones.

It just requires a soundcard to use with linphone. Other technical functionalities include DTMF (dial tones) support though RFC2833 and ENUM support (to use SIP numbers instead of SIP addresses). Linphone is a free software, released under the General Public License. Linphone is well documented and includes a SIP test server called "sipomatic" that automatically answers to calls by playing a pre-recorded message.

B. SJphone

SJphone is a dual standard VoIP soft phone, which is well-known for its good quality and interoperability. It allows the user to speak with any other soft phones running on a PC, PDA, any stand-alone IP-phone, or using ITSP (Internet

Telephony Service Provider) with any traditional wired or mobile phone. It supports both the SIP and H.323 standard sets, NAT traversal, and works with most major ITSP like Vonage, and IP-PBX and VoIP gateway vendors. SJphone is available for personal computers with Windows, MAC and Linux OS. It supports GSM, iLBC, G.711, and G729.

C. Skype

Skype is a proprietary peer-to-peer Internet telephony (VoIP) network built using Peer-to-peer (P2P) techniques and competing against established open VoIP protocols like SIP or H.323. The system has a reputation for working across different types of network connections (including firewalls and NAT) because voice packets are routed by the combined users of the free desktop software application. Skype users can speak to other Skype users for free, call traditional telephone numbers for a fee (SkypeOut), receive calls from traditional phones for a fee (SkypeIn), and receive voicemail messages for a fee. Skype uses wideband CODECS which allows it to maintain reasonable call quality at an available bandwidth of 32 Kbps. Skype has a licensed VoiceEngine product, which is a comprehensive solution that includes all of GIPS (Global IP Sound) codecs, as well as a jitter buffer, error concealment, and echo cancellation technology. The GIPS suite of codecs includes iLBC which is a fixed rate CODECS. But, as indicated on the Skype website, Skype varies its bit rate which could not happen with iLBC. The suite also contains the iSAC CODEC. According to a

GIPS engineer, the iLBC and iSAC algorithms are pretty much unrelated. iLBC is a narrowband fixed rate codec operating at 13.3 Kbps or 15.2 Kbps. The algorithm is available from IETF (RFC 3951 and 3952). iSAC is an unrelated wideband variable rate CODEC, which can adapt its operating rate between 10 Kbps and 32 Kbps. The codec is proprietary and the algorithm is unavailable. So, though Skype may indeed utilize iLBC, it must be using iSAC. As the engineer pointed out, the wideband GIPS CODECS have not been released royalty-free like iLBC. Asterisk and other open source types therefore show little interest in them. The good news is that alternatives exist - namely Speex - which supports 8, 16 and 32 kHz sample rates and is open source freeware. Skype uses TCP for signaling, and both UDP and TCP for transporting media traffic. Signaling and media traffic are not sent on the same ports.

3. EXPERIMENTAL RESULTS & ANALYSIS

Let us analyse the marked SIP packet no. 41. As the snap shot clearly explains that the network of of the ETHERTYPE. The node 102 on subnet 4 (192.168.4.102) is connected to the SIP server(192.168.7.254) which is on subnet 7. The frame network size is 608 bytes including the 4 bytes of Check Redundancy Control (CRC). The differential

time is 0608 345 since previous packet. The Destination Address comes out to be 00:23:33:22:D8:08. SIP sends request BYE from UDP port 61726 to 5060 with IP 192.168.4.102 and 192.168.7.254 respectively i.e; to SIP server.

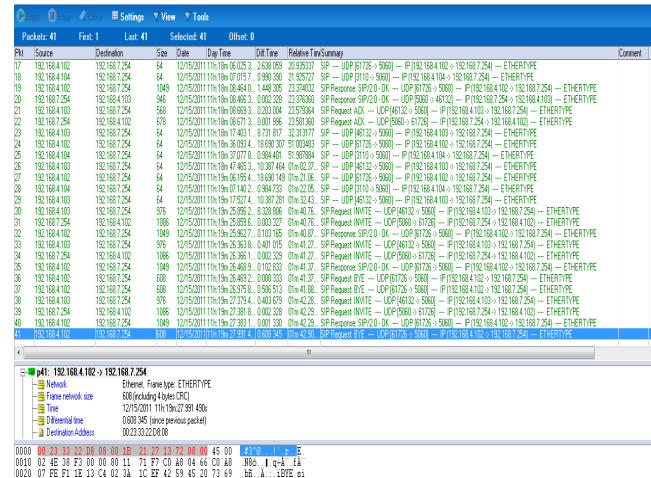


Fig. 1. SIP Packets for Wired Network Using G711 A-Law CODEC

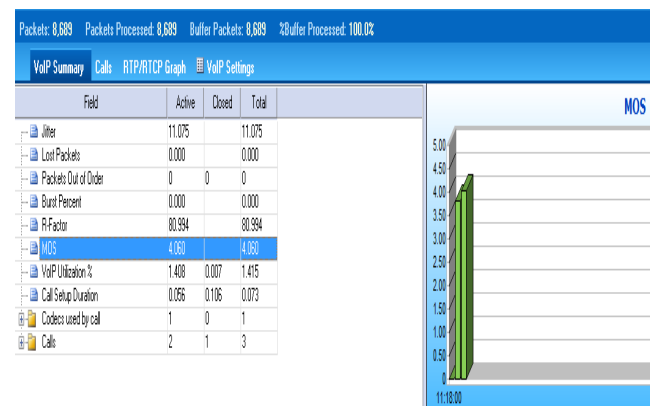


Fig. 2. MOS For Wired Network Using G 711 A-Law CODEC

As it is clearly seen that MOS value comes out to be 4.060 for the A-LAW codec which is considered to be good in terms of the network Quality of Service (QoS). Jitter is very low and the buffer processing is the cent percent which means the all the packets have been processed successfully i.e; 8689 out of 8689. VoIP Utilization of the given network is also good which comes out to be 1.408(active).

The value of jitter comes out to be 11.075 which is low and considered to be satisfied for the particular network.

The average R-factor value is 80.994, it is considered to be a satisfied value, it's likely that the network has satisfied users.

By looking at the above result, we are getting 1.408 value for the VoIP utilization%, this means that the network utilization is satisfactory.

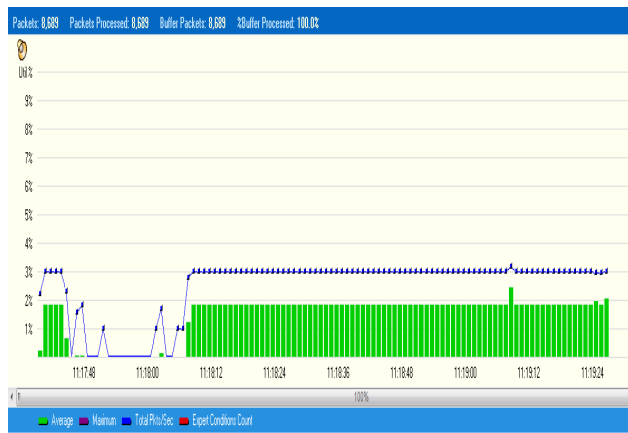


Fig. 3. Expert Summary for Wired Network Using G711 A-Law CODEC

The given picture says that for initial at the start of the voice call average and total packet processing happened but between time 11:17:48 and 11:18:06 almost no average packet processing but a little bit of total packet processing and for time interval 11:18:07 to 11:19:24, there is 100% average and total packet processing utilizing 2 and 3% respectively.

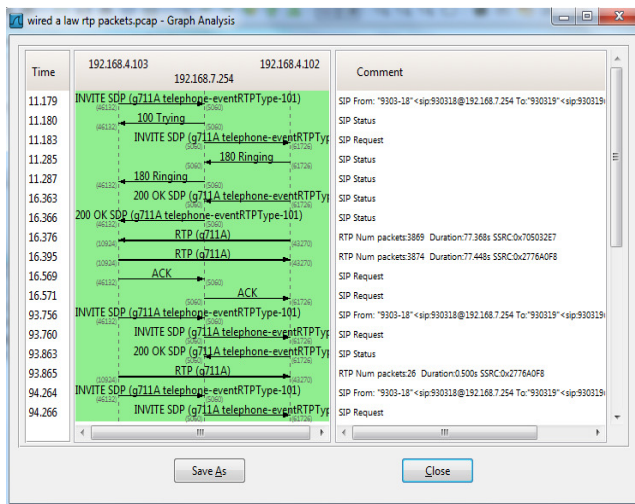


Fig. 4. Flow Graph for Wired Network Using G711 A-Law CODEC

Initially, client A (192.168.4.103) sends request to client B (192.168.4.102) through SIP server (192.168.7.254) at time 11.179 sec. SIP server then sends message 100 trying to client A which is equivalent to the phone ring at 11.180 sec. The very next moment, SIP server forwards request to client B. Client B in response sends status 180 ringing which

means it is trying to establish connection with the Client A through SIP server. SIP server forwards the same status to the client A. After a little gap at 16.363 sec., client A sends status 200 OK SDP (Session Description Protocol) to Client B through SIP server stating that CODEC g711 a law with RTP type – 101. Then, RTP (Real-time Transfer Protocol) with CODEC g711 a is established in both the directions. Then, Acknowledgement (ACK) is sent from client A to client B through SIP server.

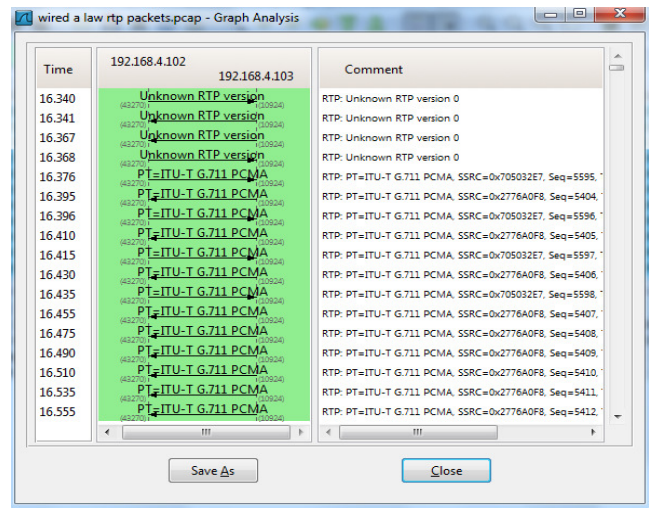


Fig.5. VoIP Call Graph Analysis for Wired Network Using G711 A-Law CODEC

Initially at the start, the version of the RTP is unknown between both the clients. Afterwards, at 16.376 sec. as it is clear from the above picture, media format ITU- T G.711 PCMA is established from client A to client B with SSRC value of 0*705032E7 and sequence 5595. Then, at the next moment i.e; at 16.395 sec. a similar media format ITU- T G.711 PCMA is established from client B to client A with SSRC value of 0*2776A0F8 and sequence 5404.

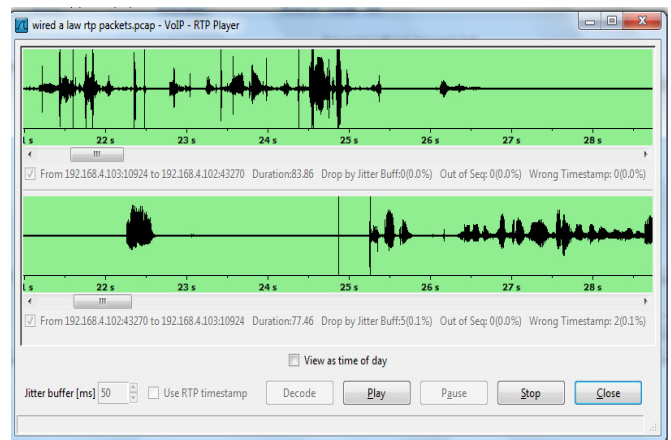


Fig. 6. Decoder for Wired Network Using G711 A-Law CODEC

As clearly seen from the above screen shot, peak or dark areas on the RTP player are speech parts in a bidirectional conversation between clients A(192.168.4.103) and clients B (192.168.4.102) and horizontal parts or without zig-zag area is silence parts. On the VAD (Voice Activity Detector) the value of speech is considered '1' and silence '0'.

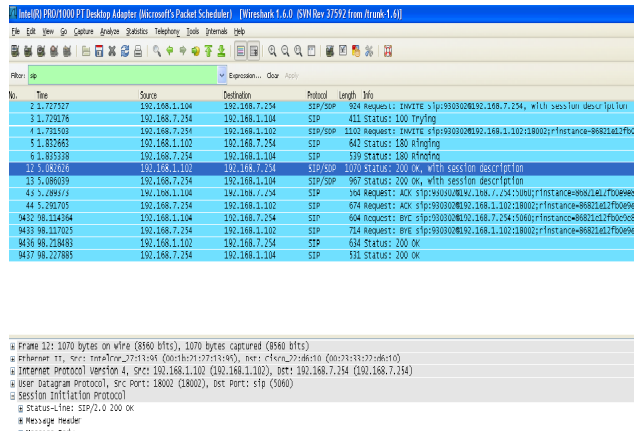


Fig. 7. SIP Packets for Wired Network Using Broadvoice CODEC

Let us analyse SIP frame no. 12 at time 5.082626 sec., which moved from source i.e; (192.168.1.102) to destination i.e; SIP server (192.168.7.254) using SIP protocol. The length of the frame is 1070 bytes.

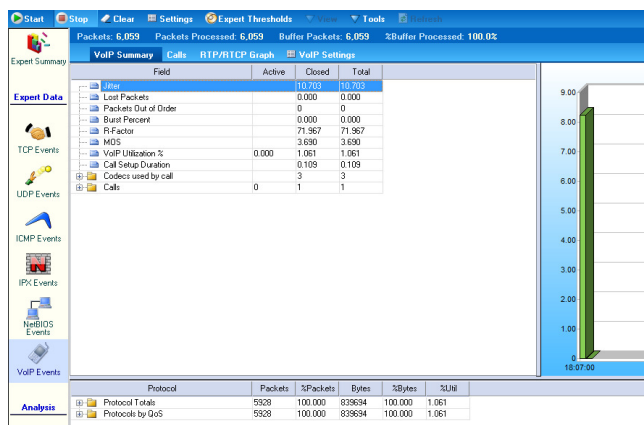


Fig. 8. MOS for Wired Network Using Broadvoice CODEC

Mean Opinion Score (MOS) is considered excellent for 5. If the average MOS falls below 3.5 it's likely that the network will have more than a few dissatisfied users. The MOS value for wired Broadvoice codec comes out to be 3.690 which is considered satisfactory with some dissatisfied users. The value of R-factor is 82.570 in this case, which considered to be satisfied for the network. Jitter measures the variability of delay in packet arrival times. Jitter in it is 11.024 to be precise which implies that the network is less affected by the external factors like noise. The VoIP utilization% using

broadvoice codec comes out to be very low 0.15 to be precise which is not considered satisfactory.

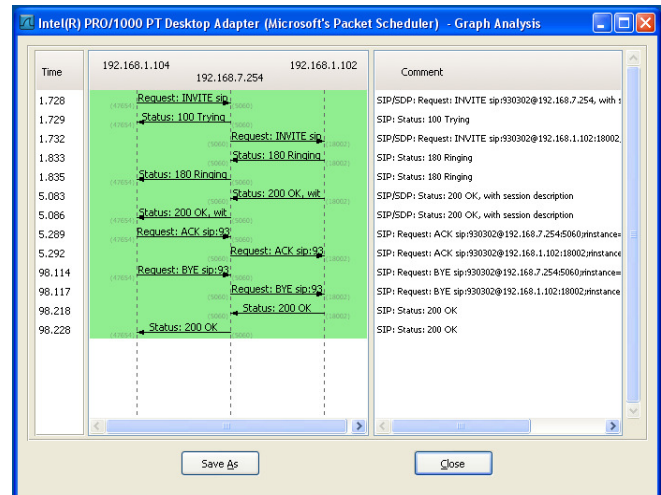
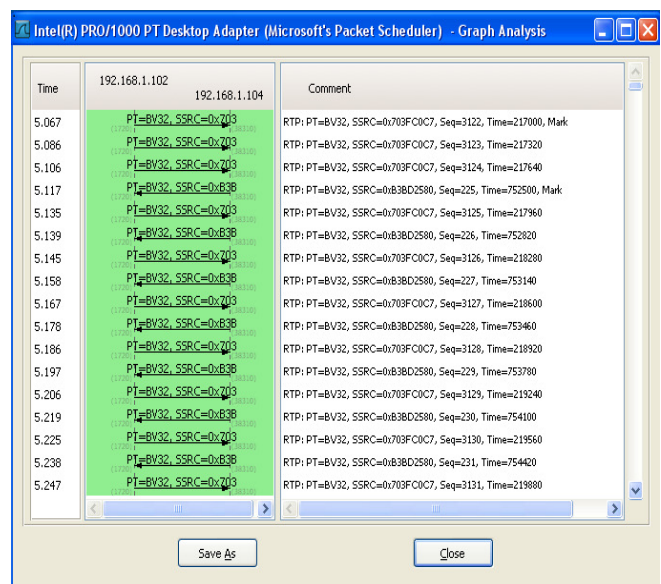


Fig. 9. Flow Graph for Wired Network Using Broadvoice CODEC

At the time 1.728 seconds the node A 192.168.1.104 sends an INVITE to the SIP server and in response at 1.729 seconds SIP server shows status 100 trying which is similar to phone ring. Then at time 1.732 seconds SIP server sends the request to node B 192.168.1.102 and at time 1.833 seconds the node B sends status ringing to the SIP server and SIP forwards it to the node A and so on.



the message from Client B to A has payload type (PT) BV32,SSRC=0x838,sequence=225 and so on....which shows pattern how VoIP call are made between both the nodes.

4. CONCLUSION

Voice over IP (VoIP) technology provides the capability of transporting voice calls across the same Ethernet office infrastructure used for data and wireless technology. The steps involved in originating a VoIP telephone call are signaling and media channel setup, digitization of the analog voice signal, encoding, packetisation, and transmission as the Internet Protocol (IP) packets over a packet-switched network. On the receiving side, similar steps (usually in the reverse order) such as reception of the IP packets, decoding of the packets and digital-to-analog conversion reproduce the original voice stream. VoIP works on Packet Switching. There are factors which determine and explain the VoIP network performance and quality are choice of CODECs, QoS, bandwidth, VoIP utilization, MOS, Jitter, R-Factor.

Having analysed and compared lab results of MOS, JITTER and R-FACTOR for CODECs G711 A -LAW & BROADVOICE, it is concluded that G 711a law produces more efficient and satisfied results for VoIP over wired network. In case of VoIP over wired network with G711 A law provides better network performance because of good MOS= 4.060 and R-FACTOR= 80.994 values than VoIP over wired network with broadvoice as a CODEC because of average values of MOS= 3.690 and R-FACTOR= 71.967.

5. FUTURE WORK

In future, CMOS RO with different number of stages at particular frequency of oscillations can be designed with great efficiency and high performance for varied applications in the areas of wireless communications and semiconductor. The study can be extended for many stages, considering analysis of all types of Ring Oscillators. Ring oscillators can also be used to measure the effects of voltage and temperature on a chip.

Ring oscillator finds applications in various fields. Thus, it can be used efficiently to make voltage controlled oscillator

in most phase locked loops (PLLs). A ring oscillator is often used to demonstrate a new hardware technology. Many wafers include a ring oscillator as part of the scribe line test structures. They can be used during wafer testing to measure the effects of manufacturing process variations. Ring oscillators can also be used to measure the effects of voltage and temperature on a chip. This way Ring Oscillator finds a number of applications in wireless communication technology and semiconductor industry.

REFERENCES

- [1] Jeong, Yeonsik, et al. (2009). VoIP over Wi-Fi Networks: Performance Analysis. Springer, 524-525.
- [2] Kim, Kyungtae And Choi, Young-June (2011). Performance Comparison of Various VoIP Codecs, ACM.
- [3] Vlaschos, Michail; Anagnostopoulos, Aris; Verscheure, Olivier; Yu, Philip S.; (4 January 2008). Online pairing of VoIP conversations. Springer-Verlag 2007, 80-81.
- [4] Luciani, L.; Gallo, A.; (November 1998). RFC2443. <http://www.irt.org/rfc/rfc2443.htm>
- [5] Schulzrinne, H.; Casner, S.; Frederick, R.; Jacobson, V.; (July 2003). RFC-3550. [online]. <http://www.ietf.org/rfc/rfc3550.txt>
- [6] A. D. Keromytis, "Voice over IP: Risks, Threats and Vulnerabilities," in Proc. Cyber Infrastructure Protection (CIP) Conference, June 2009.
- [7] A. Madhosingh, "The Design of a Differentiated SIP to Control VoIP Spam," Masters Thesis Report SPIT, CAPTCHA, Florida State University, Computer Science Department, 2006.
- [8] Network Instruments White Paper October 2007). http://www.netinst.com/assets/pdf/voip_wp.pdf
- [9] Y. Rebahi and D. Sisalem, "SIP Service Providers and the Spam Problem," in Proc. 2nd VoIP Security Workshop, June 2005.
- [10] G. Ormazabal, S. Nagpal, E. Yardeni, and H. Schulzrinne, "Secure SIP: a scalable prevention mechanism for DoS attacks on SIP based VoIP systems," in Proc. 2008 Principles, Systems and Applications of IP Telecommunications pp. 107–132.
- [11] J. Fielder, T. Gupta, S. Ehlert, T. Magedanz, and D. Sisalem, "VoIP Defender: highly scalable SIP-based security architecture," in Proc. 2007 International Conference on Principles, Systems and Applications of IP Telecommunications, pp. 11–17.

An Analysis of Unicast Routing Protocols in Mobile Adhoc Networks (MANETS)

Inhas Ashraf¹, Shabir A.Sofi², Sheikh Obaid Ahmad³, Bilal Ahmad Yatoo⁴

^{1,2,3,4}Department of Information Technology,
National Institute of Technology, Srinagar, 190006
inhas17@gmail.com

Abstract: A Mobile Ad hoc Network (MANET) is a kind of wireless ad-hoc network, and is a self-configuring network of mobile routers (and associated hosts) connected by wireless links – the union of which forms an arbitrary topology. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the larger Internet. In this paper we have analysed protocols based on transmission control approach and load distribution approach. In this paper we have analysed three important routing protocols in MANETS i.e. AODV, DSDV and DSR and later compared them based on delay, packet loss and bit rate.

1. INTRODUCTION

Wireless networks have become increasingly popular in the network industry. They can provide mobile users with ubiquitous communication capability and information access regardless of locations. Conventional wireless networks are often connected to a wired network so that the ATM (Asynchronous Transfer Mode) or Internet connections can be extended to mobile users. This kind of wireless network requires a fixed wire line backbone infrastructure. This kind of network is called Mobile Ad hoc Network (MANET).

If there are only two nodes that want to communicate with each other and are located very closely to each other, then no specific routing protocols or routing decisions are necessary. On the other hand, if there are a number of mobile hosts wishing to communicate, then the routing protocols come into play because in this case, some critical decisions have to be made such as which is the optimal route from the source to the destination which is very important because often, the mobile nodes operate on some kind of battery power. Thus it becomes necessary to transfer the data with the minimal delay so as to waste less power. There may also be some kind of compression involved which could be provided by the protocol so as to waste less bandwidth. Further, there is also a need of some type of encryption so as to protect the data from prying eyes. In addition to this, Quality of Service support is also needed so that the least packet drop can be obtained.

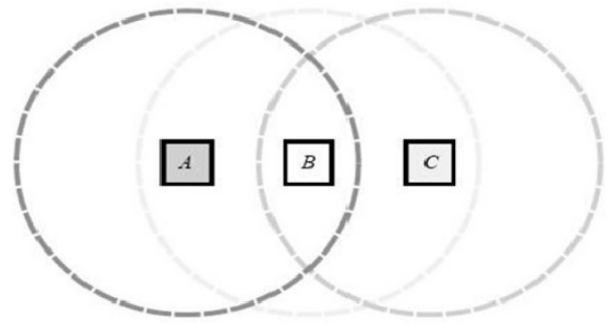


Fig 1. A mobile adhoc network with three nodes

2. ROUTING & ROUTING PROTOCOL:

Routing is the process of selecting paths in a network along which to send network traffic. Routing is performed for many kinds of networks, including the telephone network (Circuit switching), electronic data networks (such as the Internet), and transportation networks. In packet switching networks, routing directs packet forwarding, the transit of logically addressed packets from their source toward their ultimate destination through intermediate nodes, typically hardware devices called routers, bridges, gateways, firewalls, or switches

A routing protocol is a mechanism by which user traffic is directed and transported through the network from the source node to the destination node. Objectives include maximizing network performance from the application point of view application requirements while minimizing the cost of network itself in accordance with its capacity. Routing protocols can be grouped into two major classes, non-adaptive & adaptive. Non-adaptive don't base their routing decisions on measurements or estimates of the current traffic and topology. Instead, the choice of the route to choose is computed in advance, off-line, and downloaded to the routers when the network is booted. This procedure is sometimes called as static routing. Adaptive algorithms, in contrast, change their routing decisions to reflect changes in the topology, and usually the traffic as well. Adaptive algorithms differ in where they get their information. These

routing protocols help in path generation, path selection, data forwarding in addition to path maintenance between any two nodes in the network.

3. ROUTING IN CONVENTIONAL NETWORKS

The conventional routing protocols cannot be used in MANETS because these protocols are designed keeping in mind the static topology configuration of wired networks. These topologies unlike the wireless topologies do not change with time. Therefore the routing protocol for these networks do not need to control certain parameters which are associated with the change in topology such as assigning new IP addresses to the nodes in the network if some node leaves the network, finding new routes to the destination, resuming data downloads which are in progress, providing some kind of compression facility etc. Link state and distance vector would probably work very well in an ad hoc network with low mobility i.e. a network where the topology is not changing very often. Then problem that still remains is that link-state and distance vector are highly dependent on periodic control messages. As the number of network nodes can be large, the potential number of destinations is also large. This requires large and frequent exchange of data among the network nodes. This is in contradiction with the fact that all updates in a wireless interconnected ad hoc network are transmitted over the air and thus are costly in resources such as bandwidth, battery power and CPU. Because both link-state and distance vector try to maintain routes to all reachable destinations, it is necessary to maintain these routes and this also wastes resources for the same reason as above. Another characteristic for conventional protocols is that they assume bi-directional links, e.g. that the transmission between two hosts is equally well in both directions. In the wireless radio environment this is not always the case. Since most of the ad hoc routing protocols use one or more of the traditional routing algorithms as their basis, it becomes necessary to have a look at the basic functioning of these conventional routing algorithms like Link State, Distance vector and source routing.[1][3][4]

4. LINK STATE

A link-state routing protocol is one of the two main classes of routing protocols used in packet switching networks for computer communication). Examples of link-state routing protocols include OSPF and IS-IS. The link-state protocol is performed by every switching node in the network. The basic concept of link-state routing is that every node constructs a map of the connectivity to the network, in the form of a graph, showing which nodes are connected to which other nodes. Each node then independently calculates the next best logical path from it to every possible destination in the network. The collection of best paths will then form the node's routing table.

In link-state routing, each node maintains a view of the complete topology with a cost for each link. To keep these costs consistent, each node periodically broadcasts the link costs of its outgoing links to all other nodes using flooding. As each node receives this information, it updates its view of the network and applies a shortest path algorithm to choose the next-hop for each destination. Some link costs in a node view can be incorrect because of long propagation delays, partitioned networks etc. Such inconsistent network topology views can lead to formation of routing-loops. These loops are however short-lived, because they disappear in the time it takes a message to traverse the diameter of the network. This method is more reliable, easier to debug and less bandwidth-intensive than Distance Vector. It is also more complex and more compute- and memory-intensive. [5][11]

5. DISTANCE VECTOR

In distance vector, each node only monitors the cost of its outgoing links, but instead of broadcasting this information to all nodes, it periodically broadcasts to each of its neighbours an estimate of the shortest distance to every other node in the network. The receiving nodes then use this information to recalculate the routing tables, by using a shortest path algorithm. A distance-vector routing protocol requires that a router informs its neighbours of topology changes periodically and, in some cases, when a change is detected in the topology of a network. Compared to link state protocols, which require a router to inform all the nodes in a network of topology changes, distance-vector routing protocols have less computational complexity and message overhead. [5][11]

6. ROUTING PROTOCOLS IN MANETS

The routing protocols in MANETS can be classified as:-

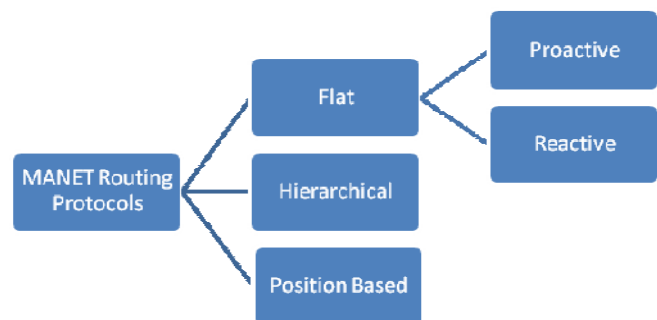


Fig. 2. Classification of MANET routing protocols

7. PROACTIVE ROUTING PROTOCOLS:

A proactive routing protocol is also called "table-driven" routing protocol. Using a proactive routing protocol, nodes in a mobile ad hoc network continuously evaluate routes to

all reachable nodes and attempt to maintain consistent, up-to-date routing information. When a network topology change occurs, respective updates must be propagated throughout the network to notify the change. Wireless Routing Protocol (WRP), the Destination Sequence Distance Vector (DSDV), and the Fisheye State Routing (FSR) are all proactive routing protocols. [10][11]

8. DSDV - THE DESTINATION SEQUENCED DISTANCE VECTOR PROTOCOL

DSDV is one of the most well known table-driven routing algorithms for MANETs. It is a distance vector protocol. In distance vector protocols, every node i maintains for each destination x a set of distances $\{d_{ij}(x)\}$ for each node j that is a neighbour of i . Node i treats neighbour k as a next hop for a packet destined to x if $d_{ik}(x)$ equals $\min_j \{d_{ij}(x)\}$. The succession of next hops chosen in this manner leads to x along the shortest path. In order to keep the distance estimates up to date, each node monitors the cost of its outgoing links and periodically broadcasts to all of its neighbours its current estimate of the shortest distance to every constantly other node in the network. The distance vector which is periodically broadcasted contains one entry for each node in the network which includes the distance from the advertising node to the destination. The distance vector algorithm described above is a classical Distributed Bellman-Ford (DBF) algorithm. DSDV is a distance vector algorithm which uses sequence numbers originated and updated by the destination, to avoid the looping problem caused by stale routing information. In DSDV, each node maintains a routing table which is and periodically updated (not on-demand) and advertised to each of the node's current neighbours. Each entry in the routing table has the last known destination sequence number. Each node periodically transmits updates, and it does so immediately when significant new information is available.

The data broadcasted by each node will contain its new sequence number and the following information for each new route: the destinations address the number of hops to reach the destination and the sequence number of the information received regarding that destination, as originally stamped by the destination. No assumptions about mobile hosts maintaining any sort of time synchronization or about the phase relationship of the update periods between the mobile nodes are made. Following the traditional distance-vector routing algorithms, these update packets contain information about which nodes are accessible from each node and the number of hops necessary to reach them. Routes with more recent sequence numbers are always the preferred basis for forwarding decisions. Of the paths with the same sequence number, those with the smallest metric (number of hops to the destination) will be used. The addresses stored in the route tables will correspond to the layer at which the DSDV protocol is operated. Operation at

layer 3 will use network layer addresses for the next hop and destination addresses, and operation at layer 2 will use layer-2 MAC addresses. [9][11] An illustration of DSDV protocol is shown below:

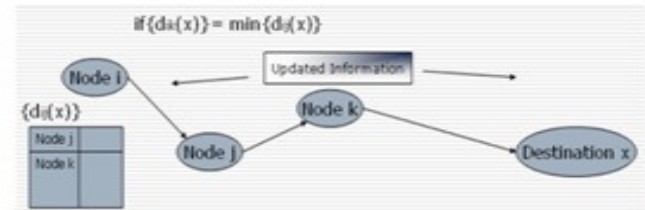


Fig. 3. An illustration of DSDV protocol

9. REACTIVE ROUTING PROTOCOLS

Reactive routing protocols for mobile ad hoc networks are also called "on-demand" routing protocols. In a reactive routing protocol, routing paths are searched only when needed. When a source node wants to send packets to the destination but no route is available, it initiates a route discovery operation. In the route discovery operation, the source broadcasts route request (RREQ) packet. When the destination or a node that has a route to the destination receives the RREQ packet, a route reply (RREP) packet is created and forwarded back to the source. Each node usually uses hello messages to notify its existence to its neighbors. Therefore, the link status to the next hop in an active route can be monitored. When a node discovers a link disconnection, it broadcasts a route error (RERR) packet to its neighbors, which in turn propagates the RERR packet towards nodes whose routes may be affected by the disconnected link. Then, the affected source can re-initiate a route discovery operation if the route is still needed. Compared to the proactive routing protocols, less control overhead is a distinct advantage of the reactive routing protocols. Thus, reactive routing protocols have better scalability than proactive routing protocols. However, when using reactive routing protocols, source nodes may suffer from long delays for route searching before they can forward data packets. Hence these protocols are not suitable for real-time applications. The Dynamic Source Routing (DSR) and Ad hoc On-demand Distance Vector routing (AODV) are examples for reactive routing protocols. [10][11]

10. AODV - THE AD HOC ON-DEMAND DISTANCE-VECTOR PROTOCOL:

AODV is another routing algorithm used in ad hoc networks. Unlike DSR, it does not use source routing, but like DSR it is on-demand. In AODV, each node maintains a routing table which is used to store destination and next hop IP addresses as well as destination sequence numbers. Each entry in the routing table has a destination address, next hop, precursor nodes list, lifetime, and distance to destination. To

initiate a route discovery process a node creates a route request (RREQ) packet. The packet contains the source node's IP address as well as the destination's IP address. The RREQ contains a broadcast ID, which is incremented each time the source node initiates a RREQ. The broadcast ID and the IP address of the source node form a unique identifier for the RREQ. The source node then broadcasts the packet and waits for a reply. When an intermediate node receives a RREQ, it checks to see if it has seen it before using the source and broadcast ID's of the packet. If it has seen the packet previously, it discards it. Otherwise it processes the RREQ packet. To process the packet the node sets up a reverse route entry for the source node in its route table which contains the ID of the neighbour through which it received the RREQ packet. In this way, the node knows how to forward a route reply packet (RREP) to the source if it receives one later. When a node receives the RREQ, it determines if indeed it is the indicated destination and, if not, if it has a route to respond to the RREQ. If either of those conditions is true, then it unicasts a route reply (RREP) message back to the source.

If both conditions are false, i.e. if it does not have a route and it is not the indicated destination, it then broadcasts the packet to its neighbours. Ultimately, the destination node will always be able to respond to the RREQ message. When an intermediate node receives the RREP, it sets up a forward path entry to the destination in its routing table. This entry contains the IP address of the destination, the IP address of the neighbour from which the RREP arrived, and the hop count or distance to the destination. After processing the RREP packet, the node forwards it toward the source. The node can later update its routing information if it discovers a better route. This could be used for QoS routing support to choose between routes based on different criteria such as reliability and delay. To provide such support additional QoS attributes would need to be created, maintained, and stored for each route in the routing table to allow the selection of the appropriate route among multiple routes to the destination [8] [11].

An illustration of AODV is shown below:

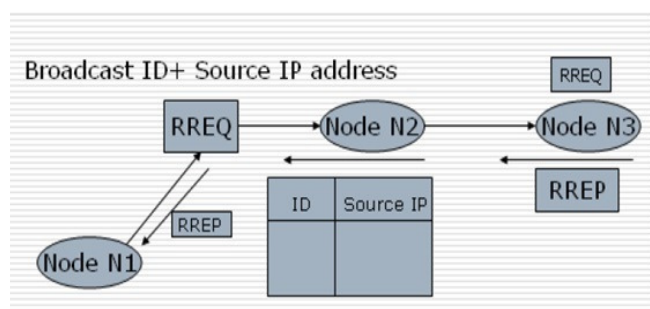


Fig. 4. An illustration of AODV protocol

11. DSR- DYNAMIC SOURCE ROUTING PROTOCOL:

DSR is one of the most well-known routing algorithms for ad hoc wireless networks. It was originally developed by Johnson, Maltz, and Broch. DSR uses source routing, which allows packet routing to be loop free. It increases its efficiency by allowing nodes that are either forwarding route discovery requests or overhearing packets through promiscuous listening mode to cache the routing information for future use. DSR is also on demand, which reduces the bandwidth use especially in situations where the mobility is low. It is a simple and efficient routing protocol for use in ad hoc networks. It has two important phases, route discovery and route maintenance. The main algorithm works in the following manner. A node that desires communication with another node first searches its route cache to see if it already has a route to the destination. If it does not, it then initiates a route discovery mechanism. This is done by sending a Route Request message. When the node gets this route request message, it searches its own cache to see if it has a route to the destination. If it does not, it then appends its id to the packet and forwards the packet to the next node; this continues until either a node with a route to the destination is encountered (i.e. has a route in its own cache) or the destination receives the packet. In that case, the node sends a route reply packet which has a list of all of the nodes that forwarded the packet to reach the destination. This constitutes the routing information needed by the source, which can then send its data packets to the destination using this newly discovered route. Although DSR can support relatively rapid rates of mobility, it is assumed that the mobility is not so high as to make flooding the only possible way to exchange packets between nodes. [2][11]

12. HIERARCHICAL ROUTING PROTOCOLS:

Typically, when wireless network size increase (beyond certain thresholds), current "flat" routing schemes become infeasible because of link and processing overhead. One way to solve this problem and to produce scalable and efficient solutions is hierarchical routing. Wireless hierarchical routing is based on the idea of organizing nodes in groups and then assigning nodes different functionalities inside and outside of a group. The Zone Routing Protocol (ZRP), Zone based Hierarchical Link State routing (ZHLS) and Hybrid Ad hoc Routing Protocol (HARP) are examples for hybrid routing protocols. [10][11]

13. SIMULATION RESULTS:

For deriving inference and results we used NS2 Simulator with it's in built patches for DSDV, AODV & DSR routing protocols and finally compared time delay, packet loss rate and bit rate by plotting the results in the form of X-graph in the same simulator.

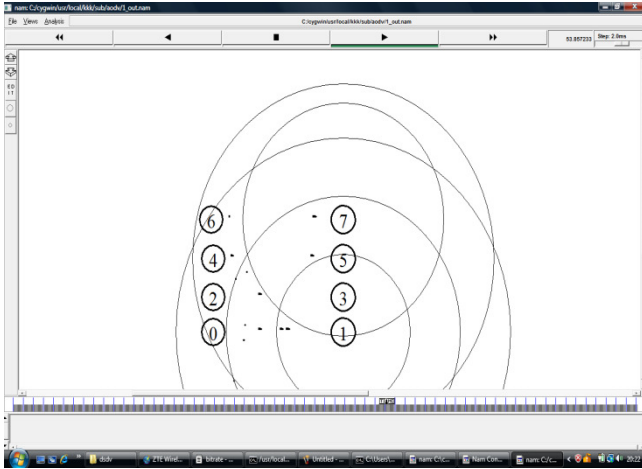


Fig. 5. Implementation of AODV protocol

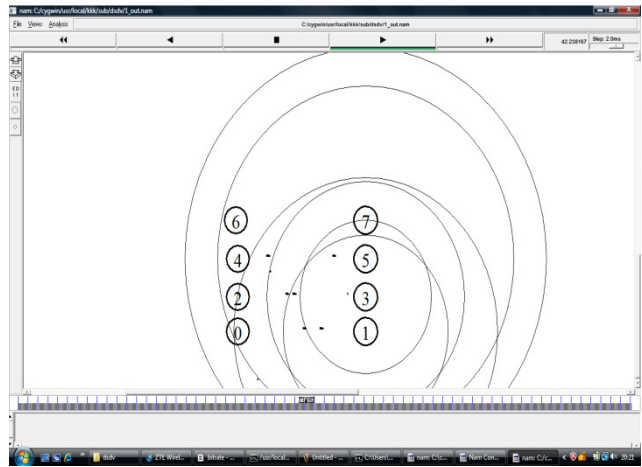


Fig. 6. Implementation of DSDV protocol

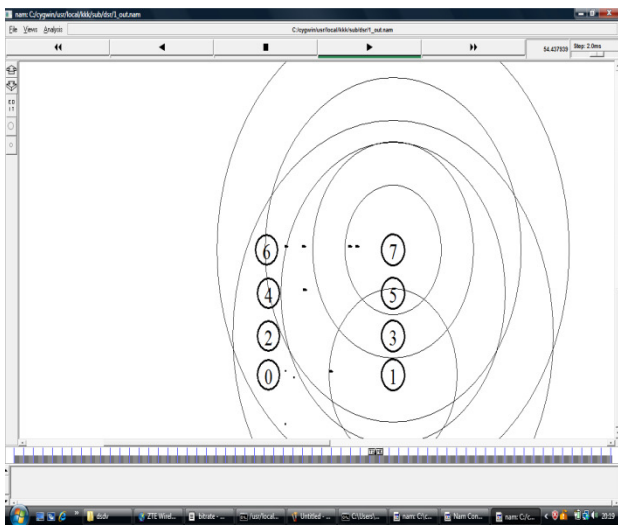


Fig. 7. Implementation of DSR protocol

14. RESULTS

COMPARISON OF DELAY

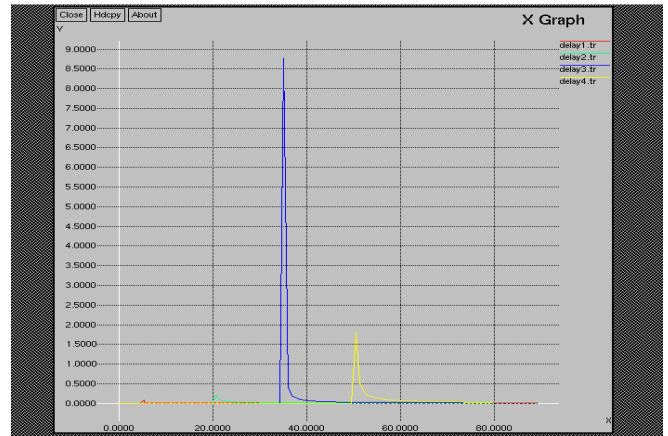


Fig. 8. Screen shot of AODV delay graph

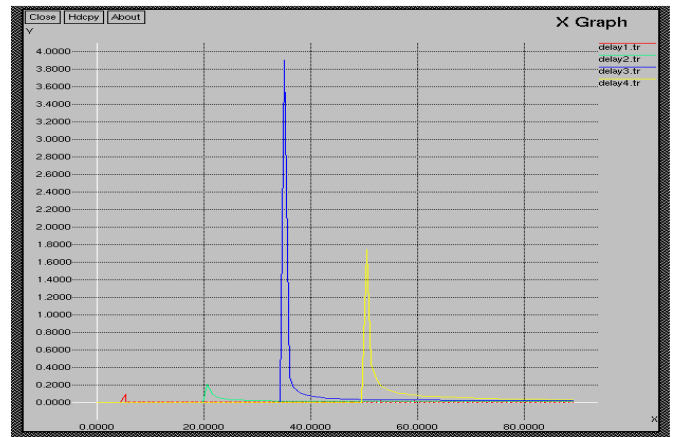


Fig. 9. Screen Shot of DSDV delay graph

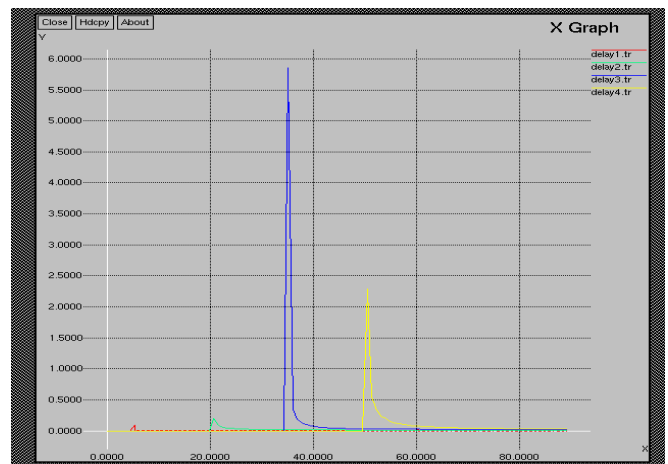


Fig. 10. Screen shot of DSR delay graph

15. INFERENCE

When the number of nodes that are sharing the network resources increase, the delay significantly increases .Furthermore the surge in delay is maximum for AODV where as it is minimum for DSDV. DSR shows an intermediate result with respect to AODV and DSDV.

16. COMPARISON OF PACKET LOSS RATE

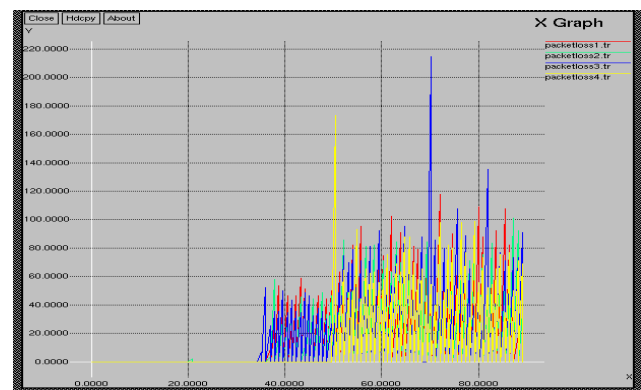


Fig. 11. Screen shot of AODV packet loss rate graph

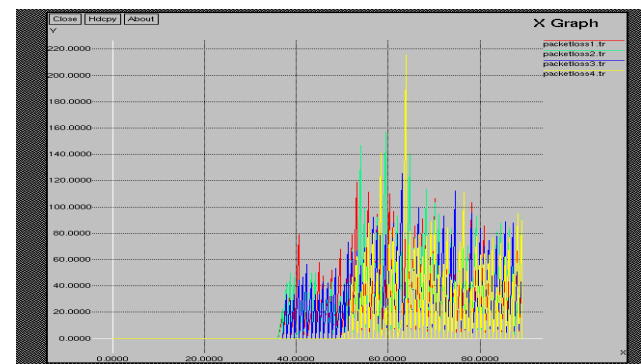


Fig 12.Screen shot of DSDV packet loss rate graph

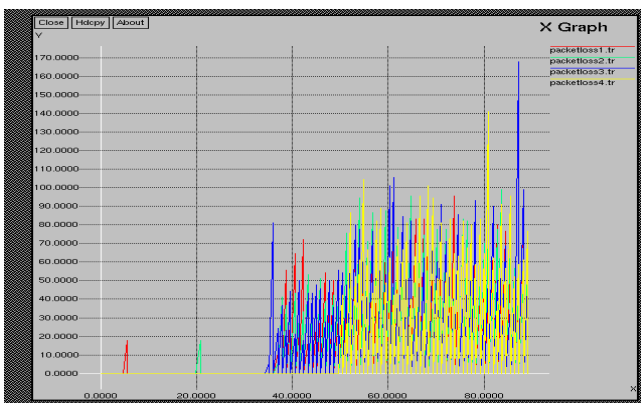


Fig. 13. Screen shot of DSR packet loss rate graph

17. INFERENCE

The figure shows a high packet drop rate whenever the number of nodes sharing network resources increases. It can be shown that the packet drop rate in the interval [5 sec, 20 sec] is 0. This can be easily justified since only one node is using the network during this time interval. However this high-quality performance is deteriorated as more nodes start sharing the network resources. Evident from the graphs the packet loss rate is highest in AODV where as the value is lesser in DSDV and DSR. However, packet loss rate is higher in DSR compared to DSDV

18. COMPARISON OF BIT RATE:

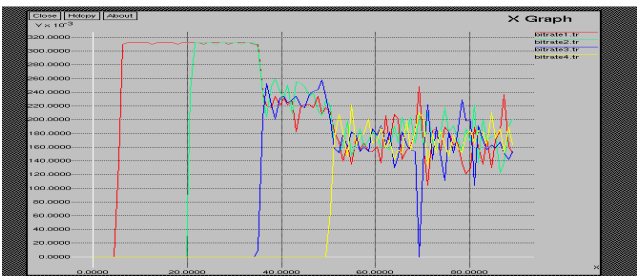


Fig. 14. Screen shot of AODV bitrate graph

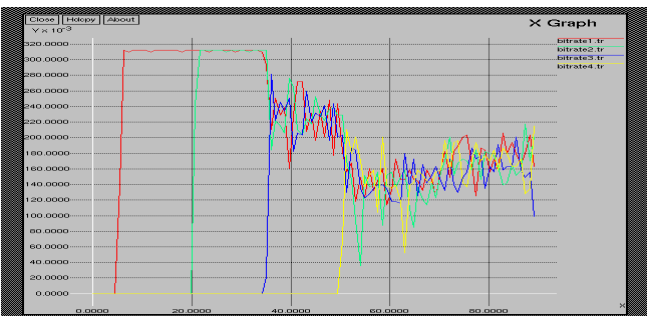


Fig. 15. Screen shot of DSDV bitrate graph

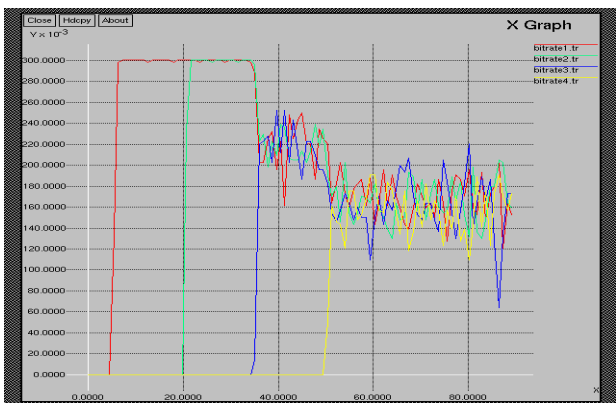


Fig. 16. Screen shot of DSR bitrate graph

19. INFERENCE

Node 1 starts transmitting at time $T = 5$ sec while Node 2 starts transmitting at time $T = 20$ sec. During the period of time [5 sec, 20 sec] Node 1 is the only transmitting node using the entire available bandwidth. This justifies the high performance of Node 1 during the specified interval of time. At time $T = 20$ sec, Node 2 starts transmission hence sharing channel resources with Node 1. This explains the heavy reduction of bit rate. In addition, the bit rate plot experiences heavier oscillations and reduction as the number of transmitting nodes increases. Oscillations are reflected in heavy disorders in network performance. In this case AODV shows the least dip in bitrate at $t = 50$ seconds followed closely by DSDV, where as DSR shows the highest dip.

20. CONCLUSION

In low mobility Ad hoc networks, for an application such as real-time streaming etc, the most appropriate routing protocol would be DSDV in comparison to DSR and AODV. Since it introduces minimum end-to-end delay, in case the number of transmitting nodes increases in the topology. If an application require minimum packet drop rate, the most appropriate protocol would be again DSDV in comparison to DSR and AODV. However if an application requires a stable bit rate i.e. with low fluctuations, the most appropriate protocol would be AODV. On the other hand, if DSR is implemented in this case, the fluctuations in bit rate will be maximum.

REFERENCES

- [1] Chenxi Zhu and M. Scott Corson. "QoS Routing for Mobile Ad Hoc Networks". In the Proc. IEEE Infocom, June 2001.
- [2] David B. Johnson and David A. Maltz. "Dynamic Source Routing in Ad Hoc Wireless Networks". In Ad Hoc Wireless Networks, Mobile Computing, T.Imielinski and H. Korth (Eds.), Chapter 5, pp 153-181, Kluwer Academic Publishers, 1996.
- [3] Andreas Tønnesen. "Mobile Ad-Hoc Networks"
- [4] Ahmed Al-Maashri and Mohamed Ould-Khaoua. "Performance Analysis of MANET Routing Protocols in the Presence of Self-Similar Traffic". IEEE, ISSN- 0742-1303, First published in Proc. of the 31st IEEE Conference on Local Computer Networks, 2006.
- [5] Elizabeth M. Royer and C-K Toh. "A Review of current Routing Protocols for Ad-hoc Mobile Wireless Networks", IEEE Personal Communications, Vol. 6, No.2, pp. 46-55, April 1999.
- [6] Sandeep Gupta. "A brief guide to ns2."
- [7] Jaroslaw Malek. "Trace graph - Network Simulator NS-2 trace files analyzer" <http://www.tracegraph.com>
- [8] C.E. Perkins, E.M. Belding-Royer and S. Das. "Ad hoc on-demand distance vector (AODV) Routing". RFC 3561, July 2003.
- [9] Charles E. Perkins and Pravin Bhagwat. "Highly Dynamic Destination- Sequenced Distance-Vector routing (DSDV) for mobile computers". In Proceedings of the SIGCOM'94 Conference on Communications Architecture, protocols and Applications, pp 234-244, August 1994.
- [10] D.P. Aggarwal and Qing-an Zeng. "Introduction to wireless and Mobile Systems". Brooks/Cole, 2005.
- [11] Handbook of wireless ad hoc networks, Mohammad Ilyas.

Review on Recent Energy Efficient Techniques in Wireless Sensor Networks

Jyoti¹, Ranjana Thalore², Manju³, Urvashi Singh⁴, M. K. Jha⁵

^{1,2,3,4,5}Mody Institute of Technology and Science, Lakshmangarh (Raj), 332311, India

¹js.ec26@gmail.com (M.Tech Student), ²cool.thalore@gmail.com (M.Tech Student)

³khurana.manju@gmail.com (Ph.D Research Scholar), ⁴urvashisingh455@gmail.com (M.Tech Student)

⁵jhaa_manish@yahoo.co.in (Professor)

Abstract: In wireless sensor network, devices or nodes are generally battery powered devices. These nodes have limited amount of initial energy that are consumed at different rates, depending on the power level. The lifetime of the network is defined as the time until the first node runs out of battery. A lot of work is going on in the area of wireless sensor networks (WSN) for energy efficient and prolonged lifetime of sensor nodes. This paper attempts to review some of the important works carried over on energy efficient design of WSN.

Keywords: Wireless sensor networks, Energy efficient, Protocol development, Routing.

1. INTRODUCTION

A wireless sensor network (WSN) consists of sensor nodes deployed over a large geographical area for monitoring physical phenomena like temperature, humidity, pressure, wind etc. Typical application areas of WSN include data collection, monitoring, surveillance, and medical telemetry and control and activation. These sensors are compact in size and besides sensing they also have limited signal processing and communication capabilities. In addition, a power source supplies the energy needed by the device to perform the programmed task. The nodes are mostly deployed in a hostile or hazardous environment, thereby making it impossible or inconvenient to replace or recharge the battery. At the same time, the sensor network should have a lifetime long enough to fulfill the application requirements. In many cases, a lifetime in order of several months or even years may be required. Therefore one of the major challenges in wireless sensor networks is to prolong the network life.

2. RELATED WORK

Various approaches have been taken to design energy efficient, energy aware WSN. Jing He *et al.* [1], has proposed conserving sensor energy and prolonging the network lifetime while guaranteeing the coverage of desired

areas or targets, called the Coverage problem. They have worked on *Activity Scheduling Problem* (ASP) [2] to

maximize the network lifetime on the premise of preserving the sensing coverage. Sensing coverage ensures that deployed sensors cover the sensing field completely [3,4] or select active sensors in a densely deployed WSN to cover the entire sensing field. The authors have formalized the target coverage problem with reliability as α -Reliable Maximum Sensor Cover (α -RMSC) problem based on failure probability concept and by using distributed algorithm. The algorithm can efficiently compute the number of α -Reliable sensor covers. The lifetime of WSN can be extended using user-defined failure probability requirements so that sensors from current active sensor cover are responsible for monitoring targets and remaining sensors are in low-power sleep mode. The proposed method can control the system's reliability easily without sacrificing the network lifetime much.

GAO De-yun *et al.* [5] proposed a simple and feasible synchronous node sleeping and waking mechanisms for small scale wireless sensor networks. They mainly focused on two mechanisms i.e. node sleep scheduling and power control. Node sleep scheduling was done by dividing all of the sensor nodes into forwarding nodes and listening nodes. Beacon frames containing sleep command from the coordinator can be forwarded to listening nodes via forwarding nodes. All the nodes in the network can enter sleep at about the same time. Through network synchronization mechanism, synchronous sleep and wake can be realized throughout the entire network. A new *power control scheme based on routing protocol* (PCBRP) in the *medium access control* (MAC) layer [6-7] is proposed. It operates with the help of routing protocol and calculates optimal transmission power according to the distance between neighbour nodes. A mapping table including optimal transmission power and node address is established during the route building procedure. The transmission power can be obtained by searching the table with the address of next-hop neighbour in subsequent data transmissions. The proposed mechanisms are implemented in sensor nodes and are evaluated in a test bed. The analysis and evaluation based on the experimental results confirmed that the

proposed energy-saving mechanisms are feasible and effective.

The unbalanced power consumption among sensor nodes may cause network partition. Chih-Yung Chang *et al.* [8] proposed efficient node placement, topology control, and MAC scheduling protocols to prolong the sensor network lifetime, balance the power consumption [9] of sensor nodes, and avoid collision. Firstly, a virtual tree topology is constructed based on Grid based WSNs. Then two node-placement techniques [10], namely Distance-based and Density-based deployment schemes, are proposed to balance the power consumption of sensor nodes. Finally, a collision-free MAC scheduling protocol is proposed to prevent the packet transmissions from collision. In addition, extension of the proposed protocols are made from a Grid-based WSN to a randomly deployed WSN, enabling the developed energy-balanced schemes to be generally applied to randomly deployed WSNs. Simulation results reveal that the developed protocols can efficiently balance each sensor node's power consumption and prolong the network lifetime in both Grid-based and randomly deployed WSNs.

Tai-Jung Chang *et al.* [11] proposed clustered-based *colour-theory-based energy efficient routing* (CEER) algorithm based on a range-free *colour-theory-based dynamic localization* algorithm, CDL [12], in which the location of a sensor node is represented as a set of RGB values. With known RGB values for each sensor node, one can find out the most possible position of a node by looking up the location database in the server. To keep track of a sensor node's location, frequently updating the RGB values of each sensor node and delivering the update to the server is necessary. However, if battery-powered nodes frequently update and report their positions, they may consume energy quickly and also waste bandwidth. The CEER selects those cluster members that are closer to the anchor than itself as next possible hops by comparing their RGB values. Among the selected cluster members, the sensor node with the highest energy level is chosen as the next hop. The CEER has no topology hole problem. Simulation results have shown that the proposed CEER algorithm can save up to 50–60% energy than *energy saving dynamic source routing* (ESDSR) [13].

Nikolaos A. Pantazis *et al.* [14] proposed a TDMA scheduling scheme for energy efficiency in order to construct an appropriate transmission schedule that achieves high levels of power conservation and at the same time reduces the end-to-end transmission time from the sensors to the gateway. Network connectivity is ensured by scheduling TDMA based wakeup intervals, which are used for propagating Wakeup messages, prior to data transmissions. The appropriate scheduling of the Wakeup intervals allows the data packets to be delayed by only one sleep interval for the end-to-end transmission from the sensors to the

appropriate gateway. This scheme was used for selecting the “optimal” end-to-end delay-limited power conservation mechanism, under various network and traffic conditions. The proposed scheme was compared to S-MAC. The choice of the S-MAC for the performance evaluation is based on the fact that S-MAC is the most well-known protocol and the majority of the schemes in the literature are compared with it. The performance evaluation shows the advantages of the proposed scheme, when both high power conservation and low delay are simultaneously desired in a static network that produces limited amounts of traffic, a situation that arises in disaster detection WSNs, where the networks are expected to retain the ability to detect and follow a rare event for a long period of time.

M. K. Jha *et al.* [15] proposed a new MAC protocol scheme, called *multi-layer MAC* (ML-MAC) protocol. In this scheme it is attempted to reduce node power consumption beyond that achieved by S-MAC and T-MAC by reducing the idle listening time and also by reducing the number of collisions. ML-MAC is a distributed contention-based MAC protocol where nodes discover their neighbours based on their radio signal level. Also, ML-MAC is a self-organizing MAC protocol that does not require a central node to control the operation of the nodes. In ML-MAC with L layers, nodes are distributed into L layers to reduce the idle listening time by a number proportional to L. The listen periods of the nodes in different layers are non-overlapping. This reduces energy consumption from two sources of energy inefficiency: idle listening and collision. ML-MAC performed well in conserving energy by having an extremely low duty cycle, reducing traffic activity at any given time, and reducing the probability of collisions.

Some routing algorithms like MST-based (*minimal spanning tree*) and LET-based (*least energy tree*) ones can reduce the total energy consumption of all the nodes in the network, they usually place too heavy burden of forwarding data on a couple of key nodes so that these nodes drain out their batteries quickly, which shortens lifetime of the WSN. Yihua Zhu *et al.* [16] proposed an *energy-efficient routing algorithm to prolong lifetime* (ERAPL) of WSN, in which a *data gathering sequence* (DGS), used to eliminate mutual transmission and loop transmission among nodes, is constructed whereby each node proportionally forwards traffic to its neighbouring node. A mathematical programming model is designed whose objective function incorporates minimal remaining energy and total energy consumption. They used *genetic algorithms* (GAs) [17] with compressed chromosome coding scheme to find the optimal solution of the proposed programming problem. The ERAPL is particularly suitable for such scenario as environment monitoring in which data are generated in a constant rate.

Nauman Aslam *et al.* [18] proposed a novel energy efficient cluster formation algorithm based on a multi-criterion

optimization technique. The technique is capable of using multiple individual metrics in the *cluster head* (CH) selection process as input while simultaneously optimizing on the energy efficiency of the individual sensor nodes as well as the overall system. Cluster formation is a process whereby sensor nodes decide with which CH they should associate among multiple choices. *Multi-criterion optimization* (MCOP) or *multi-objective decision making* is an engineering design method used in large scale complex systems to optimize the efficiency of several subsystems [19]. The motivation behind the MCOP-based cluster formation technique is to maximize network lifetime by selecting the best CH for a group of sensor nodes by considering multiple criteria such as distance of node to the CH, distance between CH and sink and residual energy.

LI Zhi-yuan *et al.* [20] has proposed a secure coverage-preserved node scheduling scheme for WSNs based on energy prediction in an uneven deployment environment. The scheme comprised of an uneven clustering algorithm based on arithmetic progression, a cover set partition algorithm based on trust and a node scheduling algorithm based on energy prediction. Simulation results showed that network lifetime of the scheme is 350 rounds longer than that of other scheduling algorithms. Furthermore, the scheme can keep a high network coverage ratio during the network lifetime and achieve the designed objective which makes energy dissipation of most nodes in WSNs balanced.

Xue Wang *et al.* [21] proposed a parallel energy-efficient coverage optimization mechanism to optimize the positions of mobile sensor nodes based on maximum entropy clustering in large-scale wireless sensor networks. According to the models of coverage and energy, stationary nodes are partitioned into clusters by *maximum entropy clustering* (MEC) [22, 23]. After identifying the boundary node of each cluster, the sensing area is divided for parallel optimization. A numerical algorithm is adopted to calculate the coverage metric of each cluster, while the lowest cost paths of the inner cluster are used to define the energy metric in which Dijkstra's algorithm [24, 25] is utilized. Then cluster heads are assigned to perform parallel particle swarm optimization to maximize the coverage metric and minimize the energy metric where a weight coefficient between the two metrics is employed to achieve a trade off between coverage area and energy efficiency. Simulations of the optimization mechanism and a target tracking application verify that coverage performance can be guaranteed by choosing a proper weight coefficient for each cluster and energy efficiency is enhanced by parallel energy-efficient optimization.

Mohamed Hefeeda *et al.* [26] has proposed a protocol, called *Probabilistic Coverage Protocol* (PCP). A key feature of this protocol is that it can be used with different sensing models, with minimal changes. Their protocol activates less sensors

than the others while maintaining the same level of coverage, consumes much less energy and extends the network lifetime. It also provides a flexible way to control the number of activated sensors versus the level of coverage achieved by the protocol.

Yingshu Li *et al.* [27] have proposed a wireless sensor-network-based scheme for monitoring composite events and delivering warnings to users in a delay-bounded and energy-efficient manner. To ensure the quality of surveillance, some applications require that if an event occurs, it needs to be detected by at least k sensors, where k is a user-defined parameter. So they examined the *Timely Energy-efficient k-Watching Event Monitoring* (TEKWEM) problem and propose a scheme, which involves an event detection model and a warning delivery model, for monitoring composite events and delivering warnings to users. The event detection model guarantees that each event can be k watched (where k is a user-defined parameter) and the warning delivery model ensures that warnings can be delivered to users timely. In this work, they take advantage of heterogeneous WSNs, where there are two kinds of nodes in a network: resource constrained sensor nodes and resource-rich gateway nodes as shown in fig.1.

Resource-constrained sensor nodes are normal sensors with limited energy, size, and weight. Gateway nodes are rich in energy and memory, have much stronger computation and communication abilities compared with sensor nodes. However, the cost of each gateway node is more expensive than that of a sensor node, therefore, fewer gateway nodes are employed in a WSN. In this work, main drawback is that only warnings are delivered once an event is detected, while event information is not recorded.

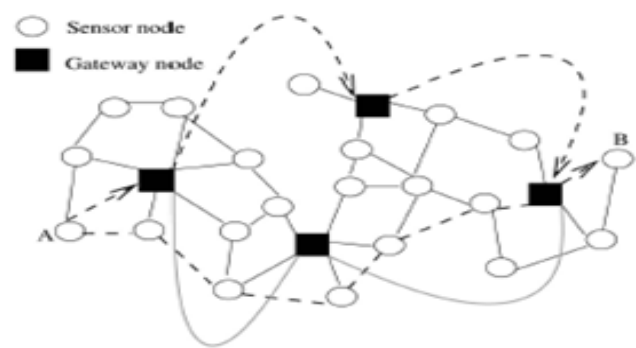


Fig. 1. A heterogeneous WSN

Nikolaos A. Pantazis *et al.* [28] has proposed a TDMA scheduling scheme for energy efficiency in order to construct an appropriate transmission schedule that achieves high levels of power conservation and at the same time reduces the end-to-end transmission time from the sensors to the gateway. The Path-Wake Up and the wake up message

aggregation strategies presented in this paper can be used for minimizing the sleep-related end-to-end delay from the sensors to the gateway and for minimizing the idle listening time, in order to decrease the consumed power for given delay levels. The proposed scheme achieves higher power conservation than other relevant schemes, when the traffic generation rate is low.

Behrouz Maham *et al.* [29] has proposed a cooperative multihop routing scheme for Rayleigh fading channels and formulated the problems of finding the minimum energy cooperative route for a wireless network under Rayleigh fading. The investigated system can achieve considerable energy savings compared to non-cooperative multihop transmission, when there is an outage probability *quality-of-service* (QoS) requirement at the destination node. Two power control schemes, i.e., *distributed* and *centralized* power allocations are derived to minimize the total transmission power given the outage probability constraint. Three efficient cooperative protocols are proposed for multihop networks that highlight a tradeoff among *energy efficiency*, *throughput*, and *complexity*, based on the topology of network.

Nikolaos A. Pantazis *et al.* [30] has focused on the WSNs energy-efficient routing techniques which are used for Health Care Communication Systems concerning the *Flat Networks Protocols* that have been developed in recent years.

Flat Networks Routing Protocols for ad hoc WSNs can be classified, according to the routing strategy, in three main different categories: *Pro-active* protocols, *Re-active* protocols and *Hybrid* protocols. According to another classification found in the literature, *Flat Networks* routing protocols for ad hoc WSNs can be categorized as *Table-driven* and *Source-initiated (Demand-driven)* respectively (*Pro-active* and *Re-active* routing protocols) as shown in fig. 2. Pure pro-active and pure re-active protocols may not fit to WSNs, Hybrid protocols combine the advantages of both pro-active and re-active routing protocols; they locally use pro-active routing and inter-locally use re-active routing.

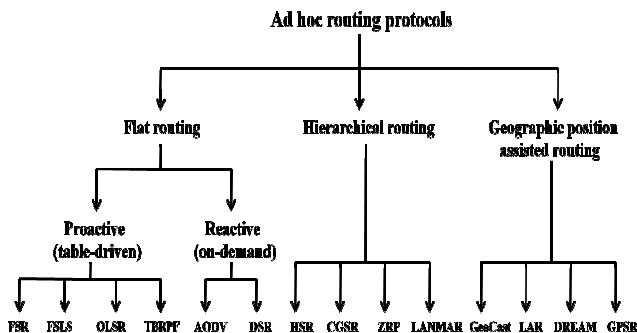


Fig. 2. Classification of Routing Protocols

Christian Domínguez-Medina *et al.* [31] have proposed *Ant Colony Optimization* (ACO). ACO based routing algorithms can add a significant contribution to assist in the maximization of the network lifetime and in the minimization of the latency in data transmissions. Two algorithms are used: The *Ant Colony Optimization Based Location Aware Routing for Wireless Sensor Networks* (ACLR) [32], and the *Energy Efficient Ant Based Routing Algorithm* (EEABR) [33]. In EEABR, that maintains small and constant ant's sizes, results in larger network lifetime. The strategy proposed in ACLR, related to the selection of the next node to jump to, which demonstrated the best option in terms of Latency.

3. CONCLUSION

In WSNs power consumption is a major issue since it determines the lifetime of a network. Several power conservation schemes have been proposed in the literature for increasing the lifetime of sensor networks and making it energy efficient, either by trying to reduce the number of transmissions through efficient routing, or by taking advantage of the sleep mode capabilities of sensor nodes. Although many techniques have been proposed in WSNs, many issues still exist and there are still many challenges that need to be solved in the sensor networks.

REFERENCES

- [1] Jing He**, Shouling Ji, Yi Pan, Yingshu Li, "Reliable and Energy Efficient Target Coverage for Wireless Sensor Networks," *TSINGHUA SCIENCE AND TECHNOLOGY ISSN 1007-0214* 02/09 pp. 464-474 Volume 16, Number 5, October 2011.
- [2] San-Yuan Wang, Kuei-Ping Shih, Yen-Da Chen, and Hsin-Hui Ku, "Preserving target area coverage in wireless sensor networks by using computational geometry," *Wireless communications and Networking Conference (WCNC), 2010 IEEE*, pp. 1-6, 18-21 April 2010.
- [3] Kar K, Banerjee S., "Node placement for connected coverage in sensor networks," *Proceedings of the First Workshop on Modeling and optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, Sophia-Antipolis, France, 2003.
- [4] Wang Y, Hu C, Tseng Y., "Efficient placement and dispatch of sensors in a wireless sensor network," *IEEE Transaction on Mobile Computing*, 2008, 7(2): 262-274.
- [5] GAO De-yun, ZHANG Lin-juan, WANG Hwang-cheng, "Energy saving with node sleep and power control mechanisms for wireless sensor networks," *The Journal of China Universities of Posts and Telecommunications* February 2011, 18(1): 49-59.
- [6] Gao T, Zheng T, Peng D, *et al.*, "A general multi-sensor node in wireless sensor networks," *Proceedings of the 2009 IEEE International Conference on Communications Technology and Applications (ICCTA'09)*, Oct 16-18, 2009, Beijing, China. Piscataway, NJ, USA: IEEE Computer Society, 2009: 406-411.

- [7] Gao D, Niu Y, Zhang H., "Micro sensor routing protocol in IPv6 wireless sensor network," *Proceedings of the 2009 IEEE International Conference on Networking, Sensing and Control (ICNSC'09)*, Mar 26–29, 2009, Okayama, Japan. Piscataway, NJ, USA: IEEE, 2009: 55–59.
- [8] Chih-Yung Chang, Hsu-Ruey Chang, "Energy-aware node placement, topology control and MAC scheduling for wireless sensor networks," *Computer Networks* 52 (2008) 2189–2204.
- [9] W. Li, C.G. Cassandras, "A minimum-power wireless sensor network self-deployment scheme," *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, March 2005, pp. 1897–1902.
- [10] P. Cheng, C.N. Chuah, X. Liu, "Energy-aware node placement in wireless sensor networks," in: *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, November 2004, pp. 3210–3214.
- [11] Tai-Jung Chang, Kuochen Wang, Yi-Ling Hsieh, "A color-theory-based energy efficient routing algorithm for mobile wireless sensor networks," *Computer Networks* 52 (2008) 531–541.
- [12] Shen-Hai Shee, Kuochen Wang, I.L. Hsieh, "Color-theory-based dynamic localization in mobile wireless sensor networks," *Proceedings of Workshop on Wireless, Ad Hoc, Sensor Networks*, August 2005.
- [13] Mohammed Tarique, Kemal E. Tepe, Mohammad Naserian, "Energy saving dynamic source routing for ad hoc wireless networks," *Proceedings of Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, April 2005, pp. 305–310.
- [14] Nikolaos A. Pantazis, Dimitrios J. Vergados, Dimitrios D. Vergados, Christos Douligeris, "Energy efficiency in wireless sensor networks using sleep mode TDMA scheduling," *Ad Hoc Networks* 7 (2009), pp. 322–343.
- [15] Manish Kumar Jha, Atul Kumar Pandey, Dipankar Pal, Anand Mohan, "An energy-efficient multi-layer MAC (ML-MAC) protocol for wireless sensor networks," *Int. J. Electron. Commun. (AEÜ)* 65 (2011), pp. 209–216.
- [16] Yi-hua Zhu, Wan-deng Wu, Jian Pan, Yi-ping Tang, "An energy-efficient data gathering algorithm to prolong lifetime of wireless sensor networks," *Computer Communications*, 33 (2010), pp. 639–647.
- [17] J.H. Holland, "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence," *MIT Press, Cambridge, MA*, 1992.
- [18] Nauman Aslam, William Phillips, William Robertson, Shyamala Sivakumar, "A multi-criterion optimization technique for energy efficient cluster formation in wireless sensor networks," *Information Fusion* 12 (2011), pp. 202–212.
- [19] W. Stadler, *Multicriteria Optimization in Engineering and in the Sciences*, Springer, 1988.
- [20] LI Zhi-yuan, WANG Ru-chuan, "Secure coverage-preserving node scheduling scheme using energy prediction for wireless sensor networks," *The Journal of China Universities of Posts and Telecommunications*, October 2010, 17(5), pp. 100–108.
- [21] Xue Wang, Junjie Ma, Sheng Wang, "Parallel energy-efficient coverage optimization with maximum entropy clustering in wireless sensor networks," *J. Parallel Distrib. Comput.* 69 (2009), pp. 838–847.
- [22] H.L. Eng, K.K. Ma, "Unsupervised image object segmentation over compressed domain," *IEEE International Conference on Image Processing*, 2000, pp. 758–761.
- [23] J. Yao, M. Dash, "Entropy-based fuzzy clustering and modeling," *Fuzzy Sets and Systems*, Volume 113, Issue 3, 1 August 2000, Pages 381–388.
- [24] M. Noto, H. Sato, "A method for the shortest path search by extended Dijkstra algorithm," *IEEE International Conference on Systems, Man, and Cybernetics*, 2000, pp. 2316–2320.
- [25] T.D. Sudhakar, "Supply restoration in distribution networks using Dijkstra's algorithm," *IEEE International Conference on Power System Technology*, 2004, pp. 640–645.
- [26] Mohamed Hefeeda, and Hossein Ahmadi, "Energy-Efficient Protocol for Deterministic and Probabilistic Coverage in Sensor Networks," *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, VOL. 21, NO. 5, pp. 579–593, MAY 2010.
- [27] Yingshu Li, Member, IEEE, Chunyu Ai, Student Member, IEEE, Chinh T. Vu, Student Member, IEEE, Yi Pan, Senior Member, IEEE, and Raheem Beyah, Senior Member, IEEE, "Delay-Bounded and Energy-Efficient Composite Event Monitoring in Heterogeneous Wireless Sensor Networks," *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, VOL. 21, NO. 9, pp. 1373–1385, SEPTEMBER 2010.
- [28] Nikolaos A. Pantazis, Dimitrios J. Vergados, Dimitrios D. Vergados, Christos Douligeris, "Energy efficiency in wireless sensor networks using sleep mode TDMA scheduling," *Ad Hoc Networks*, VOL. 7, Issue 2, pp. 322–343, March 2009.
- [29] Behrouz Maham, Member, IEEE, Are Hjørungnes, Senior Member, IEEE, and Ravi Narasimhan, Member, IEEE, "Energy-Efficient Space-Time Coded Cooperation in Outage-Restricted Multihop Wireless Networks," *IEEE TRANSACTIONS ON COMMUNICATIONS*, VOL. 59, NO. 11, pp. 3111–3121, NOVEMBER 2011.
- [30] Pantazis, N. A., Nikolidakis, Stefanos A., Vergados D, "Energy-Efficient Routing Protocols in Wireless Sensor Networks for Health Communication Systems," *PETRA'09*, June 9–13, 2009, Corfu, Greece.
- [31] Christian Domínguez-Medina and Nareli Cruz-Cortés, "Energy-Efficient and Location-Aware Ant Colony Based Routing Algorithms for Wireless Sensor Networks," *GECCO'11*, July 12–16, 2011, Dublin, Ireland.
- [32] X. Wang, Q. Li, N. Xiong, and Y. Pan., "Ant Colony Optimization-Based Location-Aware Routing for Wireless Sensor Networks," *Springer-Verlag, LNCS*, 5258: 109–120, 2008.
- [33] T. Camilo, C. Carreto, J. S. Silva, and F. Boavida, "An Energy-Efficient Ant-Based Routing Algorithm for Wireless Sensor Networks," *Springer-Verlag LNCS*, 4150: 49–59, 2006.

Study of Proactive Routing Protocol for different Buffer Size

Yashi Rajvanshi¹, Seema Rahul², Sanjay Maurya³, Mayank⁴

^{1,2,3}Student of M.Tech Digital Communication, UKTECH, Dehradun, U.K., INDIA

⁴Student of M.Tech Digital Communication

¹yashiec@gmail.com, ²seema_3485@yahoo.com,

³sanjay.skm97@gmail.com, ⁴mayankvce@gmail.com

Abstract: In Adhoc networks, nodes can communicate with each other whenever they are in communication range of each other. Nodes may be any, for e.g. Personal Digital Assistants, laptops, cell phones. Thus an ad hoc network having mobile nodes is known as Mobile Adhoc Network (MANET). In this paper we have taken the one of the Proactive Protocol Bellmann-Ford Protocol which is used as an algorithm for Distance-Vector Routing Protocols. It is best suited for short distance communication. Wireless Ad hoc are the network that doesn't require any infrastructure. Instead they are different from WLANs which require an antenna to coordinate the communication between several nodes in the network. Being infrastructure less is an important property, which leads to many applications in real world. They are useful in situations when it is not possible to build infrastructure because of less time, geographical constraints, infrastructure is destroyed or it costly to build it. MANETs are useful in rescue operations at the time of natural disasters, in military operations.

Keywords: MANET, Proactive Protocol, Bellmann-Ford, Qualnet 5.02

1. INTRODUCTION

A mobile ad hoc network is formed by mobile hosts. Some of these mobile hosts are willing to forward packets for neighbors. Examples include vehicle-to-vehicle and ship-to-ship networks that communicate with each other by relying on peer-to-peer routings. In order to ensure effective operation as the total number of nodes in the MANET becomes very large, the overhead of the employed routing algorithms should be low and independent of the total number of nodes in MANET [2].

There are mainly three types of routing protocols i.e. Reactive, Proactive and Hybrid these protocols are having different criteria for designing and classifying routing protocols for wireless ad hoc network. Ad hoc networking allows portable devices to establish communication independent of a central infrastructure. However, the fact that there is no central Infrastructure and that the devices can move randomly gives rise to various kind of problems, such as routing and security [1].

2. PROBLEM DESCRIPTION

The aim and objective of this work is to study Bellmann-Ford algorithm and to find out the stage at which buffer size we get the maximum output and at which stage we get the minimum output. So to do this we have taken different parameters such as Throughput, End-to-End Delivery of the Packets, Delay etc. This all study and Simulation Environment has been provided by one of the Wireless Sensing Network Software QUALNET 5.02.

3. WIRELESS SENSOR NETWORKS AND ROUTING PROTOCOLS IN AD HOC NETWORKS

3.1. Proactive routing protocols

In proactive routing, each node has one or more tables that contain the latest information of the routes to any node in the network. Each node maintains routing tables and responds to the changes in the network topology by propagating updates throughout the network in order to maintain a consistent view of the network. Many proactive routing protocols have been proposed, for e.g. Destination Sequence Distance Vector (DSDV), Optimized Linked State Routing (OLSR), and Bellman-Ford and so on.

3.2. Reactive protocols

Reactive routing protocols take a lazy approach to routing. They do not maintain or constantly update their route tables with the latest route topology. This type of routing creates routes only when desired by the source node. The source node initiates a process called route discovery when it requires a route to the destination. This process is completed when a route is found or when all the possible routes are examined. The process of route maintenance is carried out to maintain the established routes until either the destination becomes unavailable or when the route is no longer required. Several reactive protocols have been proposed such as Dynamic Source Routing Protocol (DSR), Ad hoc On-demand Distance Vector (AODV), Temporary Ordered Routing Algorithm (TORA), and so on.

3.3. Hybrid routing protocols

This type of protocol is combination of table-driven (Proactive) and on demand (Reactive) routing protocol i.e. it contains features of proactive as well as reactive protocol. Several hybrids routing protocols have been proposed such as Zone Routing Protocol (ZRP), Zone-based Hierarchical Link State (ZHLS) and so on, but the most popular protocol is ZRP.[3]

4. BELLMAN-FORD OVERVIEW

The algorithm known as Bellman-Ford was originally developed by Bellman [Bel58] and by Ford and Fulkerson [FF62]. It is typically described in pseudo code. [4] Bellman-Ford is used for single source shortest path along with Dijkstra Algorithm. It is a Dynamic Programming based algorithm and it work for negative weight edges. Also distributed variant of the Bellman-Ford algorithm is used in distance-vector routing protocols. The Bellman-Ford distance-vector routing algorithm is used by routers on inter networks to exchange routing information about the current status of the network and how to route packets to their destinations. The algorithm basically merges routing information provided by different routers into lookup tables. It is well defined and used on a number of popular networks. It also provides reasonable performance on small-to medium sized networks, but on larger networks the algorithm is slow at calculating updates to the network topology. In some cases, looping occurs, in which a packet goes through the same node more than once. In general, most DVR (distance-vector routing) algorithms are not suitable for larger networks that have thousands of nodes, or if the network configuration changes often. In the latter case, the routing algorithm must be able to dynamically update the routing tables quickly to accommodate changes [5]. It is used as an algorithm by distance vector routing protocols such as RIP, BGP, ISO, IDRP, NOVELL IPX. Routers that use this algorithm will maintain the distance tables, which tell the distances and shortest path to sending packets to each node in the network[6]. This protocols and algorithms currently use in the IPv4 Internet. If that protocol is used in those system of networks which have several hundreds of networks and if there is any loop formed then Bellman-ford take much time to resolve that loop so this protocol is not suitable for larger networks.

5. SIMULATION PARAMETERS

5.1 Throughput :- It is one of the dimensional parameters of the network which gives the fraction of the channel capacity used for useful transmission selects a destination at the beginning of the simulation i.e., information whether or not data packets correctly delivered to the destinations.

5.2 Average End to End delay :- The average end-to-end delay of data packets is the interval between the data packet generation time and the time when the last bit arrives at the destination [6].

5.3 Packet Delivery Sent :- It is the ratio of the number of data packets received by the CBR sink at the final destinations to the number of data packets originated by the “application layer” at the CBR sources.

5.4 Average Jitter :- Average Jitter is the variation (difference) of the inter-arrival times between the two successive packets received.

6. SIMULATION SETUP

To analyze the performance of Bellman-Ford Routing Protocol we have Simulate the different Buffer (packet) size on **Qualnet 5.02** software.

The main purpose of this simulation is to observe where we will get the Optimum information and where we will get the minimum information.

Table 1 : Simulation Parameters Values

S. No.	Parameters	Values
1	Simulator	Qualnet 5.02
2	No. of Nodes	100
3	Traffic Type	CBR
4	Terrain Area	1500 m. x 1500 m.
5	MAC Type	IEEE 802.11b
6	Antenna Type	Omni- Direction
7	Protocol	Bellmann-Ford
8	Channel Type	Wireless channel
9	Radio Propagation Model	Two-Ray Ground Model
10	Mobility Model	Random Way Point

7. RESULTS AND DISCUSSION

With the use of Qualnet 5.02 we have studied the different parameters for Bellman-Ford Protocol and their results are as follows:

7.1 Throughput (bits/sec) :- At the Transmitter side there was a throughput of 4274 bit/sec and we observed that the at 1500 buffer size we get the maximum throughput i.e. **8377** and at 10000 we will get the minimum throughput i.e. **283**.

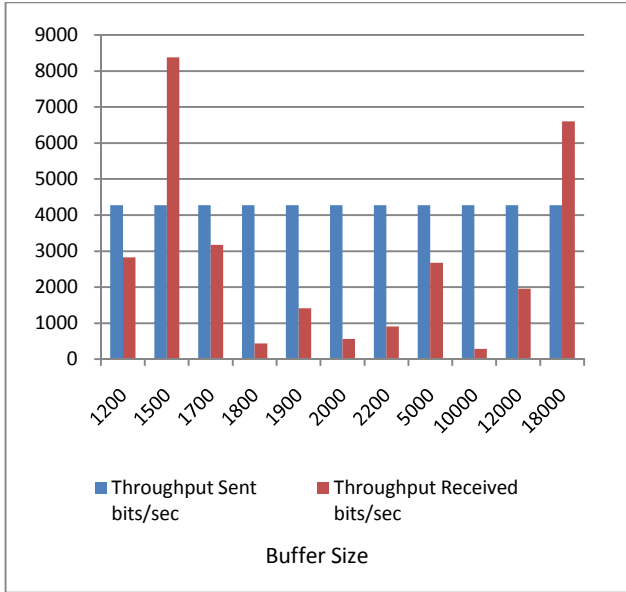


Fig. 1.1. Graph is between Throughput Vs Buffer Size

7.2 Average End to End Delay (sec) : In this it is observed that minimum delay is at 1200 buffer size i.e. **0.33 sec** and maximum delay is at 1800 buffer size i.e. **0.122 sec**.

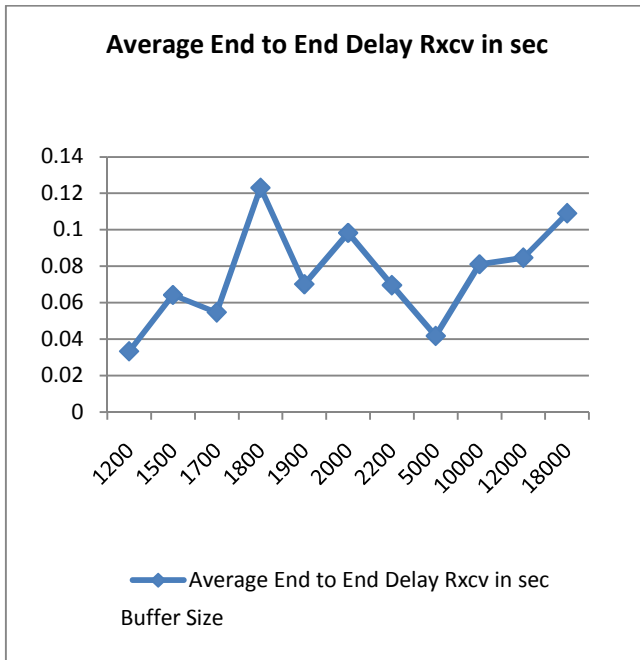


Fig 1.2:- Graph is between Average end to end delay Vs Buffer size

7.3 Packet Delivery Sent/Ratio: Initially we have sent 24 packets of data and we observe that at the buffer size of 1200 we get the minimum output i.e. only 1 packet and at the buffer size of 18000 we get more than the packet which we sent i.e. 34

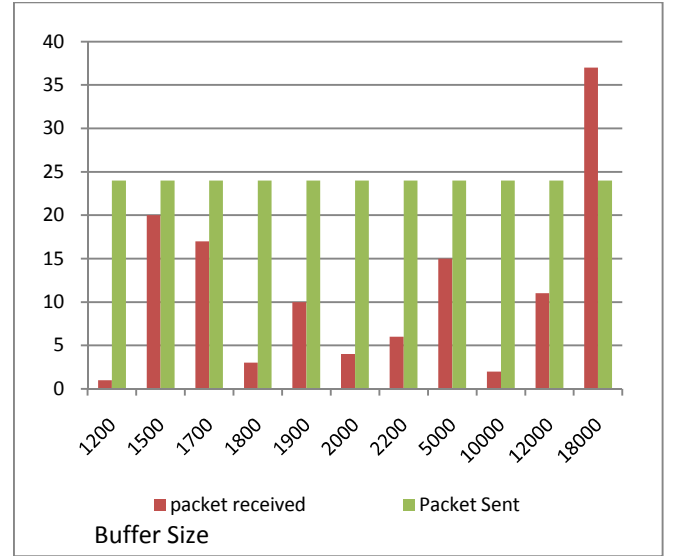


Fig. 1.3. Graph is between Packet delivery sent/ Ratio Vs Buffer size

7.4 Average Jitter (sec): The average jitter is minimum at 1200 buffer size i.e. at **0.007 sec** and maximum at a buffer size of 18000 i.e. at **0.021 sec**

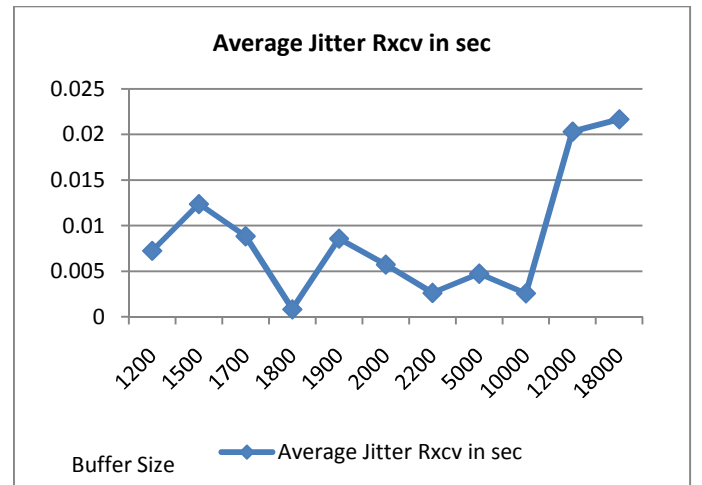


Fig. 1.4:- Graph is between Average Jitter Received Vs Buffer Size

8. CONCLUSION

In this paper we have taken the one of the Proactive Protocol which is best suited for short distance communication. Here we have taken the Scenarios of different buffer size to get the Optimize result and we find out practically also that it the best Proactive Protocol for short distance communication.

We believe that our work could be more intuitive for researchers for protocol selection and their suitability of

application in real time Scenario analysis in ad hoc networks.

REFERENCES

- [1] Kiran Salunke, Gajanan Rawalkar, Prof. S.J. Bhosale, "Implementing And Comparing DSR And DSDV Routing Protocols For Mobile Ad Hoc Networking" International Journal of Advanced Technology & Engineering Research (IJATER), VOLUME 2, ISSUE 2, MAY 2012
- [2] Yi Wang et.al, "Cluster based Location - Aware routing Protocol for Large Scale Heterogeneous MANET", in Proceeding of the Second International Multi symposium on Computer and Computational Sciences, IEEE Computer Society, 2007, pp.366-373
- [3] Mayur Tokekar and Radhika D. Joshi "Extension Of Optimized Linked State Routing Protocol for Energy Efficient Applications" International Journal on Adhoc Networking Systems (IJANS) Vol. 1, No. 2, October 2011.
- [4] Philip J. Taylor, "Specification of policy languages for network routing protocols in the Bellman-Ford family", Sept 2011,pg-
- [5] Bellman-Ford Distance-Vector Routing Algorithm (Linktionary term), 10/2012 www.linktionary.com/b/bellman.html 1/2
- [6] A. Boomarani, V.R.Sarma Dhulipala, and R M. Chandrasekaran" Throughput and Delay Comparison of MANET Routing Protocols" Int. J. Open Problems Compt. Math., Vol. 2, No. 3, September 2009 ISSN 1998-6262

Application Layer Multicast Protocols

Shubha Shukla¹, Akhilesh Kosta², Rohit Kumar³

^{1,2}*Department of Computer Science & Engineering*

Kanpur Institute of Technology, Kanpur, U.P., India shubha.shukla44@gmail.com, akhileshkosta@gmail.com

³*Department of Electronics & Communication Engineering, R.K.G.I.T. Ghaziabad, U.P., India
rohit.hbti@rediffmail.com*

Abstract: Applications such as Internet-TV and software distribution often require multicast for their delivery. Firstly IP-multicast is used for implementing multicast related functionality. However, it faces problem related to scalability, network management, deployment and support for higher layer functionality such as error, flow and congestion control. Application layer multicast is an attractive alternative solution, where multicast-related functionalities are moved to end-hosts. The key advantages Application layer multicast is it offers are flexibility, adaptability and ease of deployment. The shifting of multicast support from routers to end systems has the potential to address most problems associated with IP multicast. In last one decade, numerous algorithms and protocols have been proposed for Application Layer Multicasting. Application layer multicast, however, incur a performance penalty over router level solutions. Therefore the major concern is how to route data along the topology efficiently.

IndexTerms: Topology, Performance penalty, Scalability, overlay

1. INTRODUCTION

IP Multicast is one of the most absolute method for large bandwidth Internet applications such as video conference, IPTV, E-Learning and Telemedicine etc., But due to security and management reason IP Multicast is not enabled in Internet backbone routers. To achieve these challenges, lot of Application Layer Multicast (ALM) has been proposed.

Application layer multicast (overlay multicast) has been used for efficient point to multipoint communication across the Internet, where end systems implement all multicast related functionality including membership management and packet replication. This shifting of multicast support from routers to end systems has the potential to address most problems associated with IP multicast. However it introduces duplicate packets on physical links and incurs large end-to-end delays than IP multicast.

We specifically surveys on Application Layer Multicasting and provides much greater details about existing trends and a much deeper discussion of ALM protocols. The motivation behind studying ALM, as opposed to the other proposed alternatives to IP Multicasting, is ALM's practical success and deploys ability on today's Internet.

2. ALM PROTOCOLS

In this section we have discussed different ALM Protocols such as Narada, Delaunay Triangulation, Scribe, Nice, Overcast and OMNI. Although there are many other protocols are implemented, but we are considering only these six protocols to representing the ALM Protocols.

A. Narada Protocol

Narada is suitable for Internet conferencing applications, where participants can be both sources and receivers at the same time. In Narada, an overlay mesh is constructed by periodically adding and dropping connections between hosts. A host periodically exchanges control messages with its neighbors to maintain connectivity & build its own routing table with the shortest widest path algorithm of which bandwidth & latency are considered primary & secondary metrics, respectively.

A joining host first obtains a list of already-joined hosts from a rendezvous point (RP). The joining host then randomly selects a few of them to connect. Through periodic exchange of refresh messages among neighbors, the changes in membership due to joining or leaving of hosts can eventually be propagated to all hosts. In Narada, reverse path algorithm is used to multicast packets: when host receives a multicast packet from source, it forwards the packet to those neighbors from which host is on the shortest path to source [10].

Narada uses a periodic mesh refinement algorithm to add or drop connections so as to improve the overlay mesh. A host periodically probes some other hosts to evaluate the overall delay that can be reduced if connected to them. If the reduction of a host is above a certain threshold (the adding threshold), host connects to it. Meanwhile, host also computes the consensus cost of an edge between its neighbors. For all the shortest paths from host to the other nodes in the network, host computes the number of them include consensus cost as one of their edges. If this cost is below a certain threshold (the dropping threshold), the edge is disconnected.

Note that the adding and dropping thresholds are not fixed values, but are functions of maximum and minimum fanout of all the nodes in the overlay mesh. Therefore Narada controls the maximum and minimum fanout to avoid a host from having too many connections and becoming a bottleneck.

Narada is not scalable in term of the number of hosts. This is because [4]:

- As a flat routing algorithm, the shortest widest path algorithm holds the fact that the size of routing table is as large as the group size. Therefore the state maintained by a host is in the order of $O(N)$.
- A control message contains the states of all hosts in the group; therefore, the total control overhead in the system is high for large groups ($O(N^2)$).

Another potential problem is that Narada's convergence time can be very long. It is difficult to form a stable mesh, even after a very long time. This is due to the inconsistent criteria for adding and dropping edges.

B. Delaunay Triangulation (DT)

In DT, each host has a geographical coordinate and hosts first form an overlay mesh based on these coordinates. Compass routing is used to route a packet from one point to another. DT protocol connects the nodes together in a triangular manner so that the mesh satisfies Delaunay Triangulation property, i.e., the minimum internal angle of the adjacent triangles in the mesh are maximized [11].

In DT protocol, mesh partitioning is detected with the DT server. For each connected mesh, one and only one host is elected as leader (usually the leader is the one with the greatest coordinate). The leader periodically exchanges control messages with the server. When the mesh is partitioned, there would be more than one leaders communicating with the server. The server can then recover the partition by requesting one of them to connect to the others.

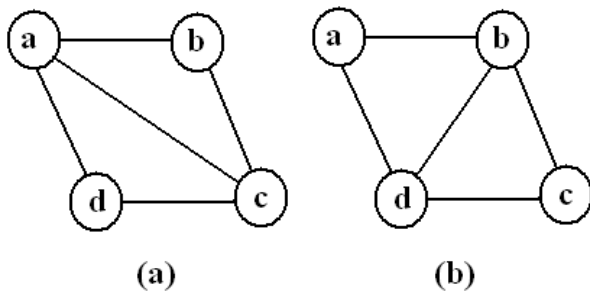


Fig. 3. Restoration of DT property by disconnecting a from c and connecting b and d.

As compared to Narada, DT protocol is much more scalable in terms of the number of hosts. This is because compass routing is a kind of local routing, which is based on the coordinates of hosts only directly connected to a node. Therefore the size of the routing table maintained by a host depends on the number of neighbors it has only, which is on average less than or equal to six. However, DT protocols come with the following two weaknesses[7]:

- *Inaccurate host location estimation* — The geographic locations of hosts in general do not correlate well with the latencies between hosts in the Internet; therefore, the end-to-end delay along a DT overlay may be quite large.
- *Single point of failure* — The partition detection and recovery scheme relies on a DT server, which forms a single point of failure.

C. Scribe Protocol

In Scribe, there are many hosts in the network but a multicast group only covers or spans a subset of them. Those hosts which are not group members take part in packet forwarding in the Scribe network. A possible application for Scribe is Internet chatroom, where usually a small set of users out of a possibly large pool belongs to the same multicast group. The large pool of *hosts* jointly takes part in forwarding packets for the group members in the system[6].

Scribe provides multicast group management for data delivery. It builds on top of Pastry, which provides the actual host-to-host routing and content-delivery mechanisms. Scribe first connects hosts together as a Pastry overlay mesh, i.e., the larger group. Then it constructs an overlay tree for each multicast group on top of the mesh such that the tree branches are embedded with the mesh edges [12]. Clearly, a host can be a tree node of multiple overlay trees. When a host receives a packet of a multicast group, it simply forwards the packet to all of its children in the corresponding multicast overlay tree. In Scribe, the non-leaf nodes are referred to as forwarders.

A multicast group in Scribe is assigned with a key (the group identifier), which is in the same key space as Node ID. A joining host first sends a join request with the group identifier as the destination key along the Pastry overlay until the request reaches a host receiving data of the same multicast group. The join request turns all hosts along the path into forwarders even though they are not members of the multicast group. Therefore, the overlay tree is an aggregation of Pastry paths from the interested hosts to the host whose key is numerically closest to the group identifier. This tree is free of loops because the distance to the destination progressively reduces upon each hop.

To maintain the connectivity of the overlay tree, each host periodically sends refresh messages to its children. To

reduce the refresh overheads, multicast packets can serve as implicit refresh messages. There is also an algorithm to remove bottleneck in the data delivery tree by limiting the children number of a host through delegation of its children to other nodes [12]. When a host is overloaded, it first identifies the multicast group which consumes the most resources and sends a control message appended with its children Node IDs to the farthest child within that group from itself. Upon receiving the control message, the child then chooses a new parent among the children listed in the message.

The size of the routing table at a host in Pastry is $O(\log_B^2 M)$, and hence Scribe is scalable in terms of group size. However, the tree-building algorithm requires a host to serve other groups not of its interest. Moreover, the performance of Pastry routes depends on the key distribution. There may be cases where even if two hosts are very close in location together, they may be separated by many hops on the Pastry overlay due to their poor match in prefixes. As a result, a high stretch value results.

D. NICE Protocol

NICE is suitable for low-bandwidth streaming applications with a large number of receivers. It organizes hosts into a multi-layer hierarchical structure, with the highest layer consists of only one host and the lowest layer consists of all the hosts in the group. A host joins a number of layers in a bottom-up manner, and hosts of the same layer are grouped into a number of clusters[8].

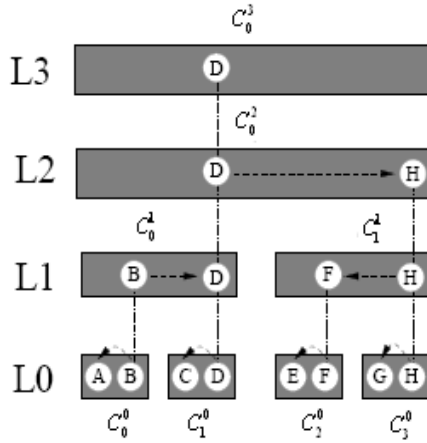


Fig. 4. In NICE, the joining host selects a cluster on layer $L0$ to join with by successively probing from highest layer to lowest layer.

We show an example of NICE in Fig-4, where shadowed boxes denote clusters, and white circles denote end-hosts. All hosts (i.e., A to H) join the bottom layer ($L0$), where hosts are grouped into a number of clusters (namely,

$C_0^0; C_0^1; C_0^2$ and C_0^3). The leader of each cluster (B, D, F and H) joins the layer one level up ($L1$). The grouping of hosts into clusters and selection of cluster leaders are then repeated.

In NICE, the clusters are organized into a hierarchical tree, with the upper cluster branching out a number of child clusters in the lower layer. A host may join multiple layers, and belong to different cluster in different layer. We denote a host's *cluster peers* as the set of all the nodes sharing clusters with the host. Unlike Scribe, the overlay tree for data delivery in NICE is not pre-constructed. When a host receives a multicast packet from a host of cluster c , it simply forwards the packet to all its cluster peers except those in cluster c . Referring Fig-4, when host D receives a multicast packet from host B of cluster C_0^1 , it forwards the packet to its cluster peers in C_0^2 and C_1^1 , i.e., nodes H and C, respectively. Therefore the maximum path length is twice the number of layers (i.e., $O(\log_k(N))$), where k is the cluster size, and the maximum node stress is equal to the product of the cluster size and the number of layers (i.e., $O(k \log_k(N))$)[7].

To maintain the overlay topology, a host periodically sends heartbeat messages to its cluster peers. Therefore, sudden leaving of any member can be detected through the loss of heartbeat messages. NICE also limits the size of a cluster from k to $3k-1$. A cluster will split into two clusters if its size is above the upper bound, or merges with another cluster if its size falls below the lower bound.

NICE is efficient in terms of end-to-end delay, since the path length of forwarding a data packet is of order $O(\log_k(N))$. However, NICE creates bottlenecks at the top-layer and higher-layer nodes, since all the joining members have to query one node at each layer of the hierarchy.

E. Overcast Protocol

Overcast is designed for single-source applications, e.g., TV-broadcasting. It tries to maximize each host's bandwidth from the source. Latency is not the major concern.

A new member joins the multicast tree by contacting its potential parents. The root node is all new nodes' default potential parent. The new node estimates its available bandwidth to this potential parent. It also estimates the bandwidth to the potential parent through each of this potential parent node's children.

If the bandwidth through any of the children approximates to the direct bandwidth to the potential parent, the closest one (in terms of network hops) of all the qualified children becomes the new potential parent and a new round

commences. If there is no qualified child, the procedure stops and the current potential parent becomes new node's parent.

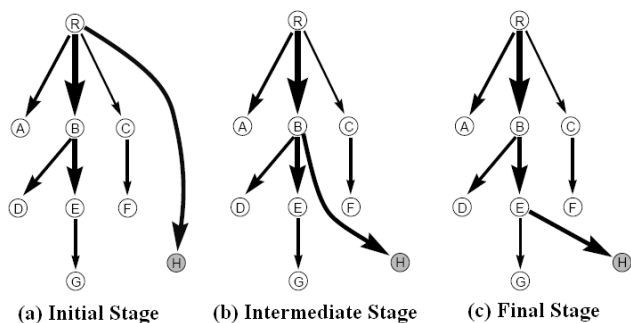


Fig. 5. The joining procedure of Overcast. Node *H* is the new member

To estimate the bandwidth, the node measures the download time of 10K bytes, which includes all the service costs. If the measured bandwidths to two nodes are within 10% of each other, we consider the two nodes equally good & select the closer one [7].

A node periodically re-evaluates its position in the tree. It measures the bandwidth to its current siblings, parent and grandparent. It will move below its sibling if that does not decrease its bandwidth back to the root. Also, it will move one level up for higher bandwidth.

In Overcast, each member maintains its ancestor list for partition avoidance and recovery. A member rejects any connection requests initiated by its ancestor(s) to avoid looping. When a member detects that its parent has left the multicast group, it connects to its ancestors one by one, from its grandparent to the root, until a live member is found. Therefore, the loading is distributed along the path to the root, and the root is not easily overloaded.

Overcast also includes an “up/down” protocol for information exchange. Each node, including the root, maintains a table of information about all its descendants and a log of all changes to the table. Each node periodically checks in with its parent. If a child fails to contact its parent within a given interval, the parent will assume the child and all its descendants have “died”. It will then modify the table. A node also modifies the table if new children arrive. During these periodical check-ins, a node reports new information that it has observed. By this protocol, the root can maintain up-to-date information about all the other nodes.

F. OMNI

This scheme allows a multicast service provider to deploy a large number of MSNs without explicit concern about

optimal placement. Once the capacity constraints of the MSNs are specified, this technique organizes them into an overlay topology, which is continuously adapted with changes in the distribution of the clients as well as changes in network conditions[13].

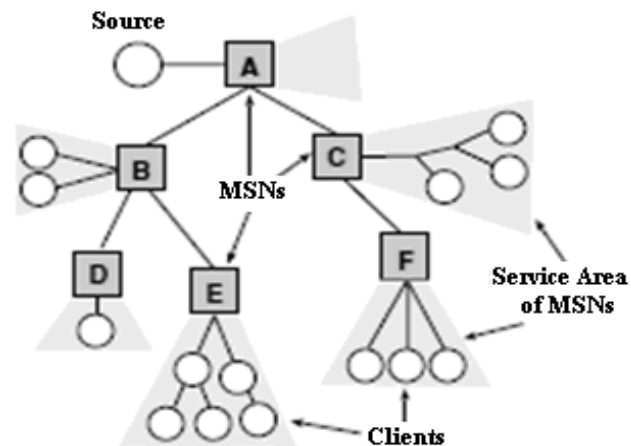


Fig. 6. OMNI Architecture

This scheme is most useful for latency-sensitive real-time applications, such as media-streaming. Media streaming applications have experienced immense popularity on the Internet. Unlike static content, real-time data cannot be pre-delivered to the different distribution points in the network.

Therefore an efficient data delivery path for real-time content is crucial for such applications. The quality of media playback typically depends on two factors: access loads experienced by the streaming server(s) and jitter experienced by the traffic on the end-to-end path. OMNI architecture addresses both these concerns as follows[13]:

- Being based on overlay architecture, it relieves the access bottleneck at the server(s), and
- By organizing the overlay to have low-latency overlay paths, it reduces the jitter at the clients.

For large scale data distributions, such as live web-casts, we assume that there is a single source. The source is connected to a single MSN, which we call the root MSN. The problem of efficient OMNI construction is as follows:

Since the goal of OMNI is to minimize the latencies to the entire client set, MSNs that serve a larger client population therefore, more important than the ones which serve only a few clients. The importance consideration of the OMNI is its ability to adapt the overlay structure based on the distribution of clients at the different MSNs. It iteratively modifies the overlay tree using localized transformations to adapt with changing distribution of MSNs, clients, as well as

network conditions. The scheme efficiently converges to near-optimal solutions.

3. CLASSIFICATION OF ALM PROTOCOLS

The classification can be done on the basis of different point such as architecture, control, approach, design objectives etc.

3.1. Architecture: Peer-to-peer or Proxi-based:

A peer-to-peer (P2P) approach constructs the overlay across end-users with all functionalities being vested with these end users. P2P architecture has the attractive property of scalability. This is because being distributed; each peer needs only to keep state for a small number of peers. Its other advantages include the simplicity of set-up and deployment, its resource sharing capability as well as dynamicity. Their vast combined resources such as physical connectivity, computing resources may be heterogeneous but they can be individually harnessed according to their capacities. P2P systems thus provide redundancy as a single failure will not radically affect the big group. Peers can be dynamically deployed in large numbers and at hot spots quickly with minimum prior configuration. The merits of using proxies as opposed to any end hosts lie mainly in their functional dedication. Being dedicated, homogeneous and better provisioned than individual hosts, proxies are thus more reliable and robust to failure. Proxies are persistent beyond the lifetime of individual hosts. They are more intelligent than end-hosts as they can provide value-added services such as being pre-configured with application specific components to render them application aware. In addition, they can be positioned at strategic positions such as co-locating with IP routers or at hotspots to provide more efficient services. However, there may be problems with the acceptance and deployment of such proxy, and when compared to P2P systems, it is less responsive to changing environment conditions as proxy placement is usually static and has to be manually deployed. ALMI, Narada, NICE and Yoid have peer-to-peer based architecture; while OMNI, Overcast are 2- tier proxy-based architecture. OMNI uses Multicast Service Nodes (MSNs) and Overcast uses Overcast nodes[13].

3.2. Control: Centralized or Distributed

A centralized approach to the overlay tree creation problem refers to the vesting of all responsibilities for group management, overlay computation and optimization with a central controller. The controller maintains group information, handles group membership, collects measurements from all members, computes optimal distribution tree and disseminates the routing tables to all members. The main advantage of a centralized controller is in the great simplification of the routing algorithms, the more efficient and simpler group management and the

provision of a reliable mechanism to prevent tree partitions and routing loops. However, the centralized nature limits its scalability and poses other reliability problem as it is more susceptible to a central point of failure while it helps to alleviate problems of tree partitions and looping. Being simple and easy to deploy, the centralized approach is a fair choice for small-scale applications[13].

The distributed approach is receiver based and it distributes the responsibilities of group membership and overlay topology computation to the individual nodes of the overlay network. It is therefore relatively more robust to failure as failure of an individual node will not impact the entire group. Having no central controller as a potential bottleneck, the distributed approach is thus more scalable and robust. However, a fully distributed approach causes excessive overheads and is not as optimal and efficient in building optimal overlay.

3.2 Approach: Mesh First or Tree First

The distributed approaches further differ in the way they create the overlay topology: some of them first create the tree topology while others first create a mesh topology. The “mesh first” approaches are Narada, Scattercast, the proposals that assign an arbitrary coordinate to each member and then performs Delaunary triangulation, Content-Addressable Networks (CAN) and Bayeux. The “tree first” approaches include YOID, Overcast, HMTP, NICE, and Tiers. Some of them (HMTP) rely on a recursive algorithm to build the tree: a newcomer first contacts the tree root, chooses the best node among the root’s children, and repeats this top-down process until it finds an appropriate parent. The clustering solutions (NICE) create a hierarchy of clusters, i.e. sets of nodes “close” to each other. Newcomers recursively cross this hierarchy to find the appropriate cluster.

3.4 Design Objective: Efficiency or Scalability

3.4.1. Source Based Trees

Narada protocol uses source based trees. In this approach, as the group size increases, control overheads increase very rapidly with the number of trees hence not efficient for large group, though suitable for a small group. [efficient but not scalable]

3.4.2. Single Shared Tree

YOID protocol uses single shared tree. It can support a fairly large group. But since all the members are in one single tree and each node has out degree bound (due to limited bandwidth and processing ability of end user terminals), the depth of the tree will be high and hence will have high latency, though suitable for non-interactive applications e.g. Video On Demand. [scalable but not efficient]

3.4.3. Multiple Shared Tree

In this approach multicast trees are more than one but far less than total number of sources. This approach is extremely useful when both the number of senders and the group size are very high. With this approach, at the one hand, total number of trees are small and hence overhead is not so large; on the other hand, tree depth is also not so large and hence latency is not so large. For n nodes and s sources, m trees are formed where $1 < m \ll s$. Thus protocol cost is only m times higher while the delay is controlled.

4. OPEN ISSUES AND FUTURE WORK

4.1. Tree Refinement

Tree refinement is the reorganized as shuffling of the nodes in the tree. This is usually conducted to enhance the system performance. In ALM, the quality of the path between any pair of members is comparable to the quality of the unicast path between that pair of members. Typically a lower diameter tree performs better than a higher diameter tree. Hence, refinement is a way to improve the quality of the ALM structure once it is already constructed. A key point is that, if a node with zero out-degree joins to a multicast session, the tree can not be extended beyond that point which ultimately increases the height of the tree. To handle such situations refinement acts as a solution. But it is an expensive operation. This is because protocols require too much information to carry out the operation. Research should therefore be conducted to find efficient mechanisms to determine whether or not refinement is applicable to a particular node.

4.2. Two Conflicting Design Goals

Another open issue is balancing the two conflicting design goals:

- Minimizing the length of the paths (usually in terms hops) to the individual destinations
- Minimizing the total number of hops to forward the packet to all the destinations

5. CONCLUSION

ALM implements multicast-related functionalities at the application level. Such technique promises to overcome the deployment problems associated with IP multicast. In ALM, since packets take more hops to reach all members, it has higher delay and stress. In this paper, we have reviewed a number of application-level protocols.

Narada, though not scalable due to its flat routing protocol, is robust in term of fault tolerance since mesh partitioning can be detected and recovered without the need of a rendezvous point. In contrast, DT is more scalable due to its

local routing protocol, although the DT server may be the single point of failure. Scribe supports applications where the tree spans only a subset of hosts. However, a host in Scribe may need to forward packets for other multicast groups, which raises some incentive issue in its deployment. In NICE, the maximum path length and node stress grows only logarithmically with the group size. Overcast targets optimal bandwidth allocation and considers latency as a supplement. OMNI adapts changing distribution MSNs, clients as well as network conditions.

Compared to IP multicasting, ALM has certain disadvantages such as longer delays and less efficient traffic generation. However, due to its overwhelming advantages for certain applications, such as immediate deployability and application-specific adaptation, it can be a practical solution to many of the existing problems in multi-user communications. The fact that an ALM protocol can be developed and deployed on the Internet without the need to make any changes to the existing network infrastructure, and the ability to evolve and apply modifications to the protocol quickly and easily at the application layer has helped the ALM approach to have a quicker start compared to other multi-user communications solutions.

The popularity of application layer multicasting continues to grow in different fields as an alternative to native IP multicasting. These include news group, video conferencing, internet games, interactive chat-lines, distant learning, and video on demand just to name a few. Although ALM is considered as an active research topic over the last decade, still there are many open issues to continue research for creating efficient and robust ALM protocols in terms of application domain requirements and the quality of service.

REFERENCES

- [1] S. E. Deering, "Multicast routing in internetworks and extended lans," *ACM SIGCOMM Computer Communication Review*, vol. 18, no. 4, pp. 55–64, Aug. 1988.
- [2] K. Sripanidkucha, A. Myers, and H. Zhng, "A third-party value-added network service approach to reliable multicast," in *Proceedings of ACM sigmetrics*, August 1999.
- [3] Sanjoy Paul, Krishan K. Sabnani, John C. Lin, and Supratik Bhattacharyya, "Reliable multicast transport protocol (RMTP)," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 3, pp. 407–421, April 1997.
- [4] Yang hua Chu, Sanjay G. Rao, Srinivasan Seshanand, and Hui Zhang, "A case for end system multicast," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, pp. 1456–1471, October 2002.
- [5] JAorg Liebeherr, Michael Nahas, and Weisheng Si, "Application-layer multicasting with Delaunay triangulation overlays," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, pp. 1472–1488, October 2002.

-
- [6] Miguel Castro, Peter Druschel, Anne-Marie Kermarec, and Antony I. T. Rowstron, "Scribe: a large-scale and decentralized applicationlevel multicast infrastructure," *IEEE Journal on Selected Areas in Communicastions*, vol. 20, no. 8, pp. 1489–1499, October 2002.
- [7] Xing Jin, Wan-Ching Wong, S.-H. Gary Chan, Hoi-Lun Ngan, "A Survey and Comparison of Application- Level Multicast Protocols," Research grant Council in Hong kong, 2006, pp. 01-22.
- [8] Suman Banerjee, Bobby Bhattacharjee, and Christopher Kommareddy, "Scalable application-layer multicast," in *ACM. Computer Communication Review*, USA, October 2002, number 4, pp. 205–217.
- [9] John Jannotti, David K. Gifford, Kirk L. Johnson, M. Frans Kasshoek, and JamesW. O'Toole, "Overcast: reliable multicasting with an overlay network," in *Proceedings of the Fourth Symposium on Operating System Design and Implementation (OSDI 2000)*. USENIX Assoc. 2000, October 2000, pp. 197–212.
- [10] Y. K. Dalal and R. M. Metcalfe, "Reverse path forwarding of broadcast packets," *Communications of the ACM*, vol. 21, no. 12, pp. 1040–1048, Dec. 1978.
- [11] Sibson R., "Locally equiangular triangulations," *Computer Journal*, vol. 3, no. 21, pp. 243–245, 1978.
- [12] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," in *Proceedings of IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, Heidelberg, Germany, November 2001, pp. 329–350.
- [13] C. K. Yeo, B. S. Lee, M. H. Er, "A survey of application level multicast techniques" *Computer communications* 27 (2004), pp. 1547-1568.
- [14] S. Banerjee, C. Kommareddy, Koushik Kar, B. Bhattacharjee, Samir Khuller, "Construction of an Efficient Overlay Multicast Infrastructure for Real-time Applications", in *IEEE journal*, vol. 2, No. 3, pp. 1521-1531, 2003.

Analysis of Energy Consumption, Throughput and Delay using Modified EQSR Routing Protocol in Wireless Sensor Network

Sunita¹, O.S. Khanna², Amandeep Kaur³

^{1,3}M.E Student (ECE), ²Associate Professor (ECE)
sunitanittr@gmail.com, oskhanna@gmail.com, aman.dhaliwal18@gmail.com
National Institute of Technical Teachers Training and Research
Sector- 26, Chandigarh, India

Abstract: In wireless sensor network, the nodes have limited battery capacity and limited initial energy. The lifetime of network is defined as the time until the first node fails. Minimization of energy and latency are of the big issues to extend the lifetime of wireless Sensor network. This paper propose a modified QoS based *Multi-path Routing* protocol (EQSR) to analyze the energy consumption, end- to- end delay and throughput in wireless sensor network and to find the optimum path, they implement Two pass Algorithm. This routing protocol provides a useful simulation platform to analyze the energy consumption, end-to-end delay and throughput in wireless sensor network.

Keywords: Wireless Sensor Networks, Energy Consumption, Multi-path Routing, quality of service, Lifetime of wireless sensor network, end- to-end delay, throughput.

1. INTRODUCTION

The exponential growth of the wireless networking is due to the fast exchange of data in such services like e-mail, internet and data transfer. The wireless network made of small or large number of tiny nodes which has limited battery power. The data routed from source to destination depending upon the battery power. Source node can easily route the data to destination, if it has sufficient battery power. If the destination node is far away from the source node, then for transmission of data, the source node should have large battery power. But after few transmissions a threshold level comes, when that source node is dead and no node is present for transmission. And the overall lifetime of network will decrease [1]. Before designing the energy-aware sensor systems, it is important to analyze the power dissipation characteristics of a wireless sensor node.

The remaining paper is organized as follows: literature is reviewed in section II. Objective of paper is discussed in section III. Methodology along with the corresponding algorithm describes in Section IV. Simulation results are discussed in Section V. Then, at last paper is concluded in section VI.

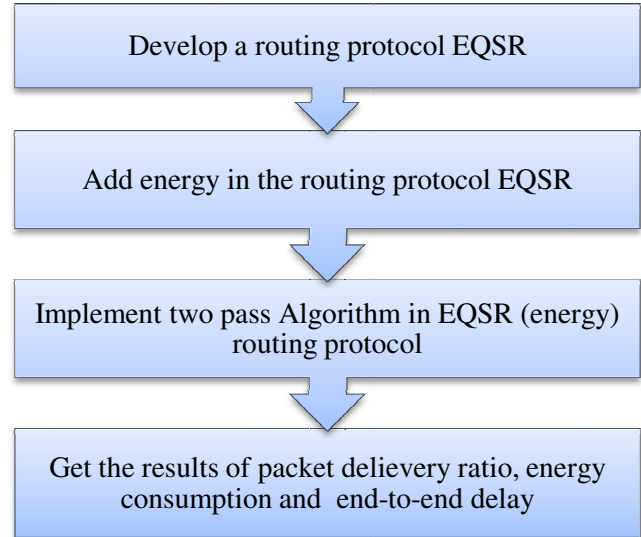
2. REVIEW OF LITERATURE

To identify the drawbacks in the system, it is necessary to analysis the power systematic [2]. S M Lambor.et.al.[3] evaluate the relationship between network lifetime and the energy consumption in a multihop wireless sensor network to enhance the performance. In evaluation, the network lifetime raised, the energy consumption decreased with the increment in number of hops and attain the minimum at critical hops. After the critical hops, the energy consumption gradually raised due to increase in cumulative energy consumption of intermediate nodes. Yash Pall et.al. [4] proposed an energy saving scheme named as maximize the lifetime of Object tracking sensor Network with node-to-node Activation Scheme (MLONAS) in which some nodes remain in sleep mode while some nodes are involved in tracking of an object. In this algorithm, when an object enter the other node's region it will activate that node and when that node start tracking the object previous one will go to sleep mode. Thus, this algorithm increases the lifetime of sensor network. Sandip Kumar Chaurasiya.et.al.[5] urged an energy-efficient routing scheme named as Enhanced Energy-Efficient Protocol with Static Clustering (E3PSC) which is the extension of an existing routing scheme, Energy-Efficient Protocol with Static Clustering (EEPSC). The proposed work divide the network into distance-based static clusters. To reduce the intra-cluster communication overhead among the nodes making the scheme more energy-efficient, cluster-head selection is performed by taking account both the spatial distribution of sensors nodes in network and their residual energy. S M Lambor.et.al.[6] analyze the effect of multiple hops on network lifetime and the energy consumption in a wireless sensor network (WSN). the results shown that as the network lifetime increases, the percentage energy consumption decreases with increase in the number of hops and gradually saturates at critical hops. Fengyuan Ren.et.al[7] proposed an Energy-Balanced Routing Protocol (EBRP) by constructing a mixed virtual potential field in terms of depth, energy density, and residual energy. The main approach is to force packets to

move toward the sink through the dense energy area so as to protect the nodes with relatively low residual energy to detect and eliminate loops; enhanced mechanisms are proposed to address the routing loop problem. Giuseppe Campobello.et.al[8] proposed a new forwarding algorithm for WSNs based on a simple splitting procedure able to increase the network lifetime. The forwarding technique is based on the Chinese Remainder Theorem and exhibits very good results in terms of energy efficiency and complexity. They also discussed the trade-off conditions between energy consumption and reliability of a novel forwarding technique for WSNs, based on the Chinese Remainder Theorem (CRT). Haibo Zhang.et.al[9] investigated the problem of maximizing network lifetime through balancing energy consumption for uniformly deployed data-gathering sensor networks. They formulate the energy consumption balancing problem as an optimal transmitting data distribution problem by combining the ideas of corona-based network division and mixed-routing strategy together with data aggregation. They first propose a localized zone-based routing scheme that guarantees balanced energy consumption among nodes within each corona. They then design an offline centralized algorithm to solve the transmitting data distribution problem aimed at balancing energy consumption among nodes in different coronas Shio Kumar Singh et.al.[10] proposed Homogenous Clustering Algorithm for wireless sensor networks to improve the energy efficiency and scalability of wireless sensor network. In this algorithm, firstly the sensor nodes are randomly clustered and then to balance the size of the clusters, they conduct self adaptive optimization. The algorithm is divided into rounds. At each round, the current cluster heads selects cluster member's node as the next cluster head. The rotation of cluster head is transparent to other cluster members. Yash Pall et.al. [11] proposed an energy saving scheme named as maximize the lifetime of Object tracking sensor Network with node-to-node Activation Scheme (MLONAS) in which some nodes remain in sleep mode while some nodes are involved in tracking of an object. In this algorithm, when an object enter the other node's region it will activate that node and when that node start tracking the object previous one will go to sleep mode. Thus, this algorithm increases the lifetime of sensor network.

3. OBJECTIVE OF PAPER

The previous work has been done on energy consumption in Energy Efficient and QoS multipath aware routing (EQSR), to increase the lifetime of the network. The author proposes service differentiation. The problem in previous paper was the energy consumption was increased. The end-to-end delay also increased. Due to which the loss of packets increases and the throughput increased [15]. So, the present work focus on these three parameters; energy consumption, end-to-end delay and packet delivery ratio. The main steps of objective of present work are:



4. METHODOLOGY

This section describes the methodology of present work to improve the lifetime of the network. Methodology includes Assumptions, Sensor Network Model, and EQSR protocol. The main methodology is shown in Fig 1.

Assumptions

Assume that there are 200 number of sensor nodes are randomly deployed in the wireless sensor network of dimension 1000m* 1000m and having average distance between them is d. All sensors have same transmission range of T. the parameter metrics are energy consumption E, throughput G, and end-to-end delay D. The parameters of network are shown in Table I.

Table 1: Parameters of Network

Parameter of Network	Description
T	Transmission range of sensors
d	Distance between source and sink
E	Energy consumption
G	Throughput
D	End –to-end delay

Sensor Network Model

The sensor nodes are randomly scattered in a sensor field as shown in fig. Each of these scattered sensor nodes have the capabilities to sense, collect information and routed back to the sink in fig. Through internet or satellite, the sink may communicate with the task manager node. There are three main components of a sensor network; sensor nodes, sink and sensed events as shown in Fig 2. The main assumption made by the most of the network architectures that sensor nodes are static. If the nodes are dynamic, then the message routed from or to moving nodes becomes difficult. The

sensed events can be either static or dynamic. Monitoring static events make the network in a reactive mode.

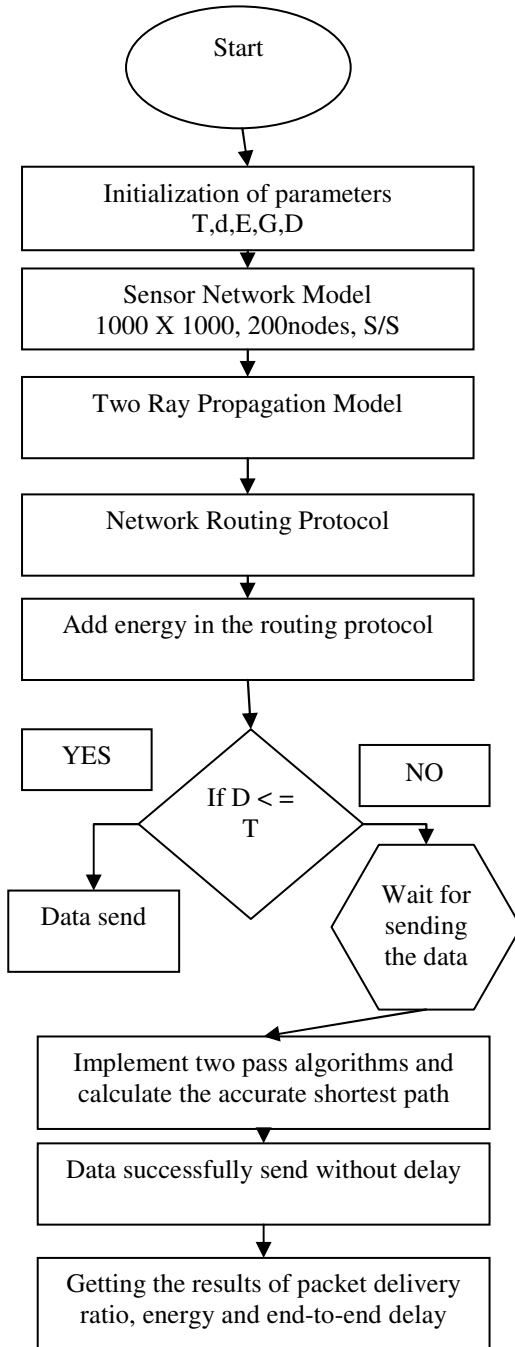
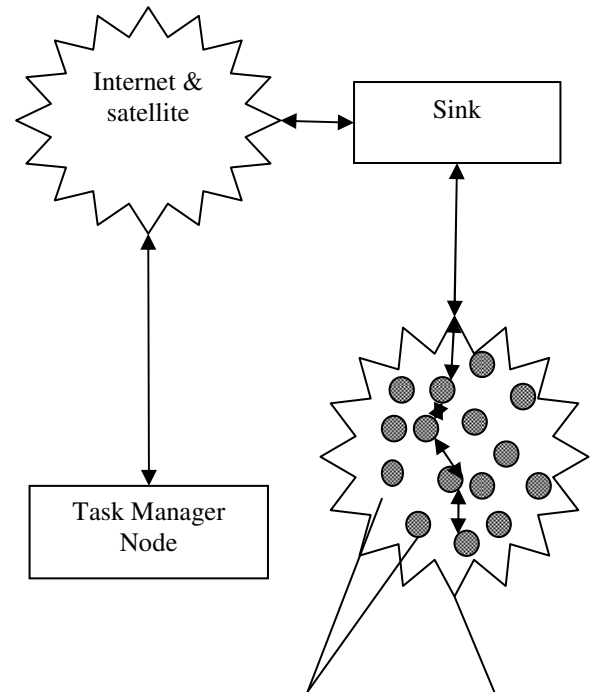


Fig. 1. Methodology for Energy Efficient Protocol for WSN

To simplify the network model, a few assumptions [1] that are taken as follows:

- Number of sensors n are uniformly scattered within a square field A . The BS is placed far away from the square field A .

- After deployment, all sensors and BS are stationary. The location of BS is known by each node. Each sensor has enough energy so can communicate with BS directly.
- Depending on the what the distance of sensors nodes between receiver or base-station, the sensors nodes can varied the transmit power by using the power control.
- If the transmission power known to the sensor nodes, then they can find the approximate distance based on the received signal strength.
- All sensor nodes are homogenous and location unaware.



Sensor nodes Sensor Network

Fig. 2. Sensor Network Model

Channel Propagation Model

To evaluate the performance of communication systems, a propagation channel should be accurate. For designing the communication system, the propagation modeling is most tough task. Signal propagation models are used to find the mean strength of the distance between transmitter and receiver. There are many factors affecting the electromagnetic propagation are categorized into three parts; reflection, diffraction and scattering [13]. Depending upon the distance between the transmitter and the receiver, there are two types of propagation models; free space model and the multipath fading model. If the distance between the transmitter and the receiver is less than the cross-over distance (d_1), the friss free space model is used (d^2

attenuation) , and if the distance is greater than d_1 ,then the two- ray ground propagation model is used (d^4 attenuation) [13]. In proposed work, two ray ground propagation models is used and its equation is as follow,

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4}$$

Where

$P_r(d)$ is the received power given at transmitter-receiver separation of d ,

P_t is the transmit power,

G_t is the gain of the transmitting antenna,

G_r is the gain of the receiving antenna,

h_r is the height of the receiving antenna above ground,

h_t is the height of the transmitting antenna above ground, and

d is the distance between the transmitter and the receiver.

If the interference comes in the path of the signal to be received, the signal is attenuated as d^4 .

EQSR Protocol (Energy Efficient and QoS multipath aware routing)

Routing protocols are the set of rules, according which each sensor has to play some roles like collecting the data from the neighboring nodes and transmit it to the sink. This protocol includes the multipath routing and QoS. Multipath routing establishes the multipath paths between source and sink. The main characteristics of multipath routing are load balancing and reliability. If we send multiple copies of data over different paths, it allows resilience to failure of number of paths [14]. The routing algorithm includes the shortest path through which the source sends data to sink. Successfully delivery of packets/data to the sink, mainly depend on three factors; transmission range, distance between source and sink, cost function. Often, large amount of packets are generated at the source side and if the sink node is in its transmission range, then the packets will send. If the sink node is not in transmission range of the source node, then the packets will not send and packets will accumulated and got congested. And the quality information is not reached at the destination.

Let T be the transmission range of sensor node. If A is source and B is sink, then if direct A sends data to B, but B is very far or away from its transmission range.

Let's transmission range of node=25m

If distance \leq transmission range

Send data

else

Wait for other which helps in sending data

So, three forwarding nodes P, Q and R are used in the route of source and sink. Routing table will be updated according to the cost function metrics. Now if these forwarding nodes have same energy, same payload, same link quality. Then same problem arises that which path will be followed to send the data to sink.

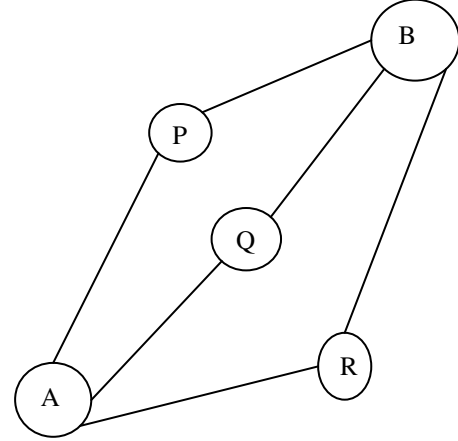


Fig. 3. Forwarding nodes

Shortest path has low cost function. To find the optimum path, we have to calculate the cost function first. Node has low cost function; data will be routed through that path. Sometimes the calculation of cost function is not accurate. But we often generally that path and data delivery failed and the energy consumption increased along with increased delay. So the proposed work focuses on the calculation of cost function. So this paper suggests Two Pass Algorithm which calculates cost function twice to get the best or optimum path.

In this algorithm, firstly find the mean and then compute the variance.

$$\text{Mean} = A = \sum_{j=1}^n \frac{x_j}{n}$$

$$\text{Variance} = \sigma^2 = \sum_{i=1}^n \frac{(x_i - A)^2}{n - 1}$$

And its pseudo code is

Two_pass_algorithim(data) :

j=0

Total 1 = 0

Total 2= 0

For j in data

j = j+1

Total 1 =Total 1+j

$$\text{Mean} = \frac{\text{Total } 1}{j}$$

For j in data

$$\text{Total } 2 = \text{Total } 2 + (x - \text{mean}) * (x - \text{mean})$$

$$\text{Variance} = \frac{\text{Total } 2}{j-1}$$

Return variance;

5. SIMULATION RESULTS

Energy consumption and end-to-end delay are the main issues in wireless sensor network. So, the proposed work focus on these two big issues. The previous work uses the routing protocol Energy Efficient and QoS multipath aware routing (EQSR) to increase the lifetime of the network. The problem in previous paper was the energy consumption was increased [15]. But the proposed work analyzes the energy consumption, end- to-end delay and throughput.

For analyze, the simulations have been done using Network Simulator version 2.34 (NS 2). NS-2 is scalable and open source used for the simulation behaviors of wired or wireless network functions and protocols [16]. The simulation network consists of 200 sensor nodes that are randomly scattered in the square field of 1000 * 1000. All nodes have same transmission range of 25m.

The simulation parameters are shown in Table 2. During the simulation, the values obtained for EQSR in terms of energy consumption and packet delivery ratio are listed in Table 3 and Table 4. From, Fig 4., it is noticed that in, energy consumption graph, the values from coordinates (0, 0) to (40, 0) are constant, that means at these points there is constant, not more and not less. Data is transmitted in proper manner and there is no much load on nodes. Then from coordinates (40, 0) to (90, 0.02), the values slightly increases. It means at these point, there is slightly increase in energy consumption. But the transmission of data gone at full fledge from source to

Table 2: Simulation parameters

Network field	1000 m × 1000 m
Propagation model	Two ray ground
Number of sensors	200
Number of sinks/number of sources	1/1
Transmission range	25m
Packet size (data + overhead)	1024 bytes
Sub-packet size	256 bytes
Transmit power	15mW
Receive power	13mW
Initial battery power	100j
MAC layer	IEEE 802.11
Simulation time	400 s

From Fig 5, it shown that in EQSR, the destination until the whole energy of network is diminished.

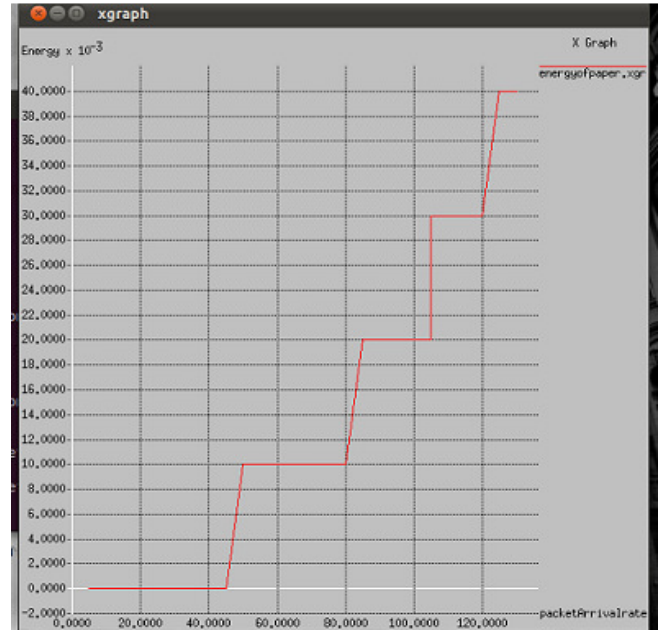


Fig. 4. Energy consumption vs packet arrival rate (packets/sec)

There is improvement in the energy consumption by using two pass algorithms and the data is routed through the optimum path. The values of packet delivery ratio are constant from coordinates (0, 1) to (30, 1), and from coordinates (30, 1) to (100, 0.98), the values slightly decreases.



Fig. 5. Packet Delivery ratio vs packet arrival rate (packets/sec)



Fig. 6. end-to-end delay vs packet arrival rate (packets/sec)

From, Fig 6., it is noticed that in EQSR, the values of end-to-end delay decreases linearly from co-ordinates (0,100) to coordinates (120, 0), due low cost function.

Table 3: Energy consumption of EQSR

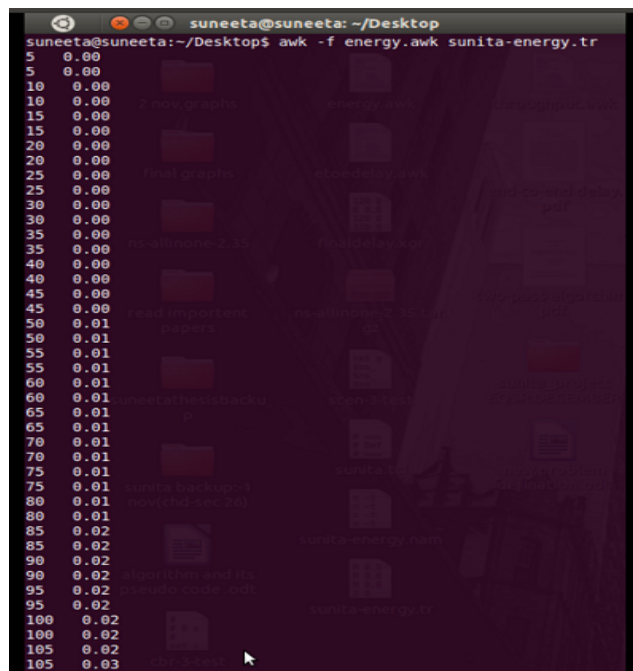
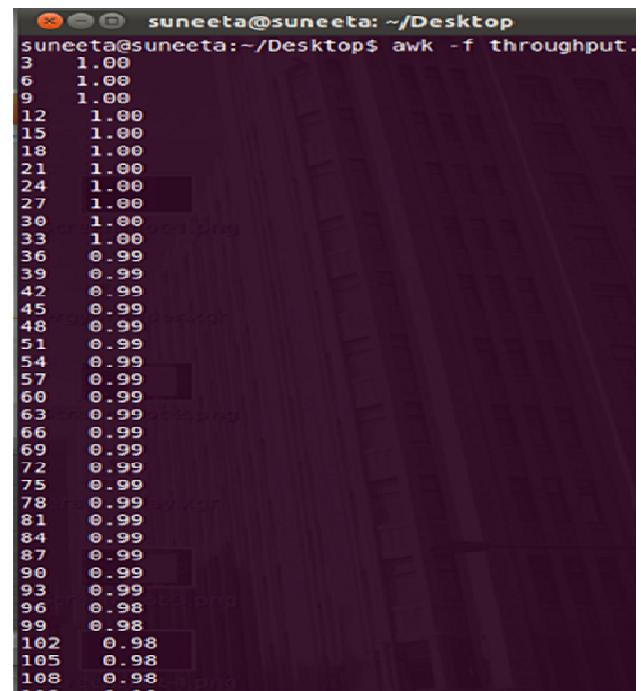


Table 4: Packet Delivery ratio of EQSR



6. CONCLUSION

This paper analyzes EQSR protocol; An Energy efficient and quality of service aware multi-path routing protocol. The performance of EQSR is analyzed using NS2 simulator. After analysis, it comes to know that EQSR has large energy savings, low end-to-end delay and high throughput. Future work can be done on the buffer size and path length.

REFERENCES

- [1] Praveen Kaushik; Jyoti Singhai, "Energy Efficient Routing Algorithm for Maximizing the Minimum Lifetime of Wireless Sensor Network: A Review", *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)* , vol.2 , no.2, 2011.
- [2] Vijay Raghunathan, Curt Schurgers, Sung Park, and Mani B. Srivasta, "Energy-Aware Wireless Microsensor Networks", *IEEE Conference on SIGNAL PROCESSING MAGAZINE*, pp. 1053-5888, 2002.
- [3] S M Lambor, S M Joshi, "Performance Analysis of Network Lifetime and Energy Consumption in a Multi-Hop Wireless Sensor Network", *2nd International Conference and workshop on Emerging Trends in Technology (ICWET)*, 2011.
- [4] Yash Pall, Lalit K. Awasthi, A. J. Singh "Maximize the Lifetime of Object Tracking Sensor Network with Node-to-Node Activation Scheme", *IEEE International Advance Computing Conference (IACC)*, 2009.
- [5] Sandip Kumar Chaurasiya, Tumpa Pal, Sipra Das Bit "An Enhanced Energy-Efficient Protocol with Static Clustering for WSN", *International Conference on Information Networking (ICOIN)*, 2011.
- [6] Fengyuan Ren, Jiao Zhang, Tao He, Chuang Lin, and Sajal K. Das, "EBRP: Energy-Balanced Routing Protocol for Data

- Gathering in Wireless Sensor Networks, IEEE Transactions On Parallel And Distributed Systems, vol. 22, 2011.
- [7] S M Lambor and S M Joshi, "Effect of Multiple Hops on Network Lifetime and Energy Consumption in a Wireless Sensor Network," *International Conference and Workshop on Emerging Trends in Technology (ICWET)*, 2011.
- [8] Giuseppe Campobello, Salvatore Serrano, Alessandro Leonardi, and Sergio Palazzo, "Trade-Offs between Energy Saving and Reliability in Low Duty Cycle Wireless Sensor Networks Using a Packet Splitting Forwarding Technique," *Journal on Wireless Communications and Networking*, 2010.
- [9] Haibo Zhang and Hong Shen, "Balancing Energy Consumption to Maximize Network Lifetime in Data-Gathering," *IEEE Transactions On Parallel And Distributed Systems*, vol. 20, 2009.
- [10] Shio Kumar Singh, M P Singh, and D K Singh, "Energy Efficient Homogenous Clustering Algorithm for Wireless Sensor Networks", *International Journal of Wireless & Mobile Networks (IJWMN)*, vol.2, 2010.
- [11] Yash Pall, Lalit K. Awasthi, A. J. Singh, "Maximize the Lifetime of Object Tracking Sensor Network with Node-to-Node Activation Scheme", *IEEE International Advance Computing Conference (IACC)*, 2009.
- [12] B. Chen, K. Jamieson, H. Balakrishnan and R. Morris, "SPAN: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks", *IEEE Conference on Wireless Networks*, vol. 8, pp. 481-494, 2002.
- [13] Taleb Moazzeni, "A Wireless Propagation Channel Model with Meteorological Quantities Using Neural Networks", *IEEE Conference on Grid and Cooperative Computing*, pp. 1 – 4, 2006.
- [14] Valera, A.; Seah, W.K.G.; Rao, S.V., "CHAMP: a highly-resilient and energy-efficient routing protocol for mobile ad hoc networks", *International Workshop on Mobile and Wireless Communications Network*, pp. 43-47, 2002.
- [15] Jalel Ben-Othman, Bashir Yahya, "Energy efficient and QoS based routing protocol for wireless sensor network", *IEEE Journals of parallel and Distributed computing*, vol. 70, pp. 849-857, 2010.
- [16] Teerawat Issariyakul, Ekram Hossain, "Introduction to Network Simulator NS2", *Springer Science and Business Media, LLC*, 2009.

Data Dissemination Protocol Based on Different Groups of Grid in WSN

Divya Sharma¹, Chandni², Kanika Sharma³

^{1,2}ME Student, Department of ECE, NITTTR, Chandigarh

³Assistant Professor, Department of ECE, NITTTR, Chandigarh

¹divya13jan@gmail.com, ²chandni.smiley08@gmail.com

Abstract: In Wireless Sensor Networks (WSNs), the nodes energy is limited, so energy consumption is essential for designing efficient routing. In this paper we proposed Data Dissemination Protocol based on different groups of Grid in WSN. The characteristic of this algorithm is to divide WSN into grids according to information of the position of nodes, and those nodes are organized within the grid by the clustering way. The clustering head is chosen according to energy level of the cluster nodes and going to make forcefully sleep the nodes which are below some predefined power up to all the nodes in that group and will improve the life time & packet reception ratio (PRR) of the sensor network.

Keywords: Data gathering, Energy-balanced routing, Grid network, Packet Reception Ratio, Wireless Sensor Network.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) is an emerging technology with a wide range of potential applications such as patient monitoring systems, earthquake detection, environment monitoring etc. Sensor networks are also being deployed also for military applications, such as navigation, surveillance, security and target tracking management [1]. A wireless sensor networks is a collection of nodes organized into a cooperative network. Each node consists of processing capability may contain multiple types of memory have an RF transceiver, have a power source and accommodate various sensors. A sensor network is consists of a large number of sensor nodes, which are densely deployed either inside the phenomenon or very close to it. Sensor networks spatially distributed autonomous sensors to monitor physical and environmental conditions at different locations, such as temperature, pressure, motion sound, vibration etc. For WSNs, many protocols have been specifically designed must be efficient, fast, resource friendly where energy awareness is an essential design issue. In wireless sensor networks, there are unique challenges with regards to unit power consumption, overall size and heat transfer. Formal verification is the process used to enable trust and security issues to be verified in relation to security protocol design for the information communications sector as shown in Fig.1.1. WSNs typically consist of small, inexpensive, resource-constrained devices that communicate among each other using a multi-hop wireless network.

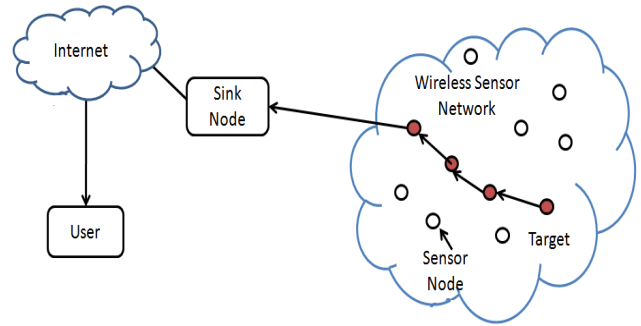


Fig.1.1 Architecture of Wireless Sensor Network

Each node, called a sensor node, has one sensor, embedded processors, limited memory, low power radio and is normally battery operated. Each sensor node of the network is responsible for sensing an event locally which is desired and at end user event is reported which is for relaying a remote event sensed by other sensor nodes. Sensor has limited energy resources and their functionality continues until their energy is finished. Therefore, applications and protocols for WSNs should be carefully designed in terms of energy-efficient manner so that the lifetime of sensor can be longer. The sensing element of a sensor probes the surrounding environment [2]. The components of sensor node are sensing unit, processing unit, transmission unit, power unit which are shown in the Fig.1.2.

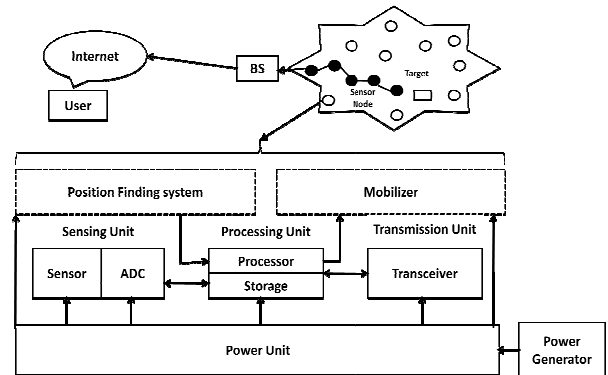


Fig. 1.2. Components of Sensor node

The power saving modes of operation are sensor nodes communicate using shortest paths, the shorter the packets, the more dominance of startup energy, operation in a power saving mode is energy efficient in that case when the time spent in that mode is more than a certain threshold.

In wireless sensor networks, the following steps can be taken to save energy caused by communication are scheduling the state of the nodes. (i.e. transmitting, receiving, idle or sleep), change the transmission range between the sensing nodes, using data collecting methods and efficient routing, as in the case of overhearing avoid the unwanted data handling. If an interesting event is detected, after performing signal processing of the observed data, sensors communicate this data to the sink or base station using a radio based link [3]. This communication happens in a single or multi-hop fashion depending on the location of the sensing node and the node has to access the medium and then transmit the data. Thus, medium access control (MAC) protocol plays an essential role in WSN. As stated earlier, these MAC protocols should be energy efficient. Because mobile nodes have limited battery power, it is therefore very important to use energy in mobile ad hoc networks (MANETs) efficiently [4].

To overcome these shortcomings, a new routing algorithm called data dissemination protocol based on different groups of grid in WSN is proposed. The characteristic of this algorithm is to divide WSN into grids according to information of the position of nodes, and those nodes are organized within the grid by the clustering way [5]. The clustering head is chose according to energy level of the cluster nodes and going to make forcefully sleep the nodes which are below some predefined power up to all the nodes in that group and will improve the life time & packet reception ratio (PRR) of the sensor network.

The rest of this paper is organized in sections as follows. In Section II surveys literature studies on Grid based energy aware routing and improve Packet Reception Ratio techniques. Section III presents our proposed protocol where we discuss the idea of data dissemination protocol based on different groups of grid in WSN. Some related performance such as PRR and energy consumption presented in section IV. Finally, Section V, paper is concluded.

2. RELATED WORK

The development and deployment of WSNs have taken traditional network topologies in new directions. Wireless sensor network topologies are Star, Tree, Circular and Grid.

a. Star Topology- Star networks are connected to a centralized communication hub (sink) and the nodes cannot communicate directly with each other. The entire communication must be routed through the centralized hub.

Each node is then a “client” while the central hub is the “server or sink” as shown in Fig.2.1. But there is disadvantage of single path communication [16].

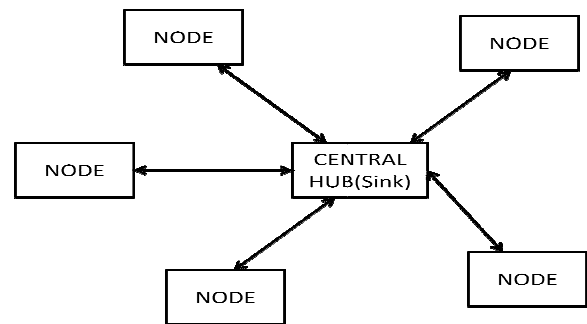


Fig. 2.1. Star Topology

b. Tree Topology: Tree network use a central hub called a root node as the main communication router. In the hierarchy, central hub is one level below from the root node. This lower level forms a star network. The tree network can be considered a hybrid of both the Star and Peer to Peer networking topologies as shown in Fig 2.2. In sensor network path may be single hop or multi hop, sensor node for getting data sense the environment and sent them to the sink and sensor forwards them to its parent after receives data messages from its children. It is important to find an optimal shortest path tree with maximum lifetime and shorter delay but slightly high time complexity and but more suitable for distributed implementation. There is problem into the load balancing scheme at each level of the fat tree and there is communication in between two nodes [12].

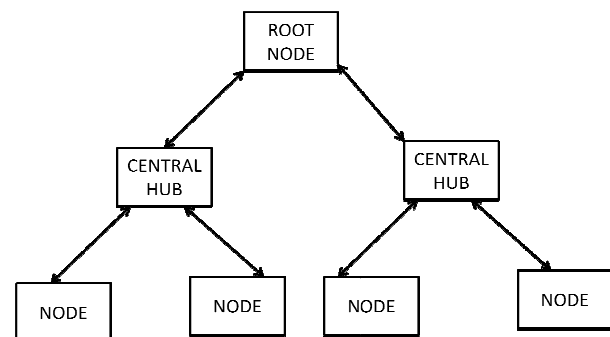


Fig. 2.2. Tree Topology

c. Circular Topology: In this topology, there is a circular sensing area and that the sensing area has a sink (at center). The sensor nodes sense the event of interest and transmit these data to the sink. The nodes are randomly deployed with uniform density all around the sink as shown in Fig.2.3. Depending on the distance of a node from the sink and the transmission range of the nodes, data have to traverse single or multiple hops before being received by the sink. At particular time slot only some of the nodes are participating

for communication, whereas other nodes are in sleep mode. So, it conserves energy. Due to the various time slots, number of collisions at the sink also has been reduced. Circular topology of this routing algorithm has number of Tiers (Tier1, Tier2,). The node which is on the diagonal, follow its original path for communication. Each of these nodes has two possible paths for routing. Depending on the energy level of the path, it selects the path and forwards the packets. The circular web topology is easy to establish, easy to maintain, and more efficient.[7]

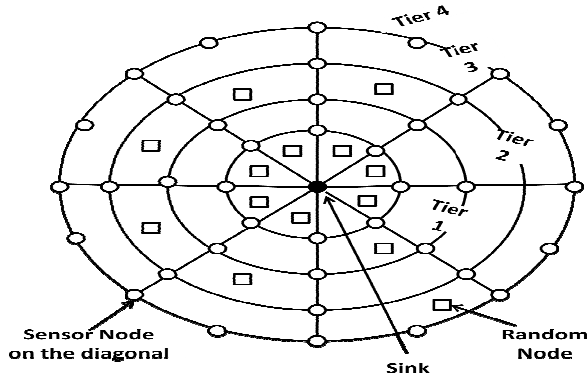


Fig. 2.3. Circular Topology

d. Grid Topology- The sensor network field dividing into grids as shown in Fig 2.4. In sensor network, inside each grid, one node is selected as a master node which is responsible for delivering the data generated by any node in that grid and for routing the data received from other master nodes in the neighbor grids of the sensor network. For each master node, multiple paths that connect the master node to the sink Original path for routing is taken as the diagonal paths between the sink and the master node. Grid-based multi-path routing protocol intended to route packets fast, utilize and extend sensor nodes energy in addition to avoiding and handling network congestion when happens.[13]

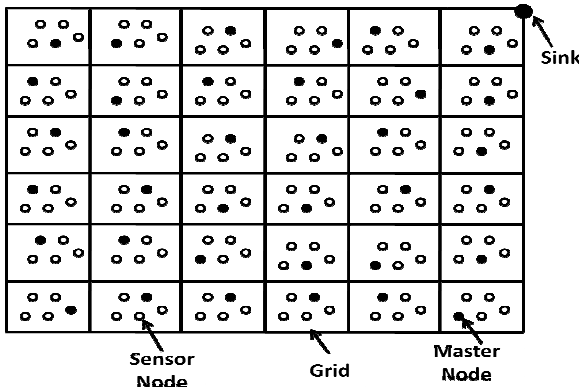


Fig. 2.4 Grid Topology

Several protocols have been proposed in the literature to address the problem of routing delay sensitive data in wireless sensor networks. The objective of these protocols is to route the packets before missing their deadlines with minimum energy consumption. In this section, we discuss related work that handles multi-path routing and congestion control issues in WSNs. From literature survey observed that in the previous section various routing techniques and topologies have been proposed to overcome the problem of delay, congestion in Wireless Sensor Networks. Every technique has its own advantages and disadvantages depending upon the applications. Sensors can be placed anywhere in home, environment etc. to collect the information as sensors have limited battery that is why energy consumption is more important in WSNs.

Network lifetime predictably is an essential system requirement for the type of WSN used in safety critical and highly reliable applications. In sensor network, at any time interval all sensor nodes are critical WSNs in these times should meet the lifetime constraint, else it may cause effects that involve losses that can be economic or even fatalities. In earlier works researches do met with the condition that the sensors which are not in use will go in sleep mode and with respect to that there will lead in the total network battery life [6]. Highly energy constrained environment of Wireless Sensor Network demands the maximum possible energy conservation that can achieve long life time of network. The proposed work focused on "Improving the lifetime of routing protocol and increase the packet reception ratio in WSNs". In a large scale WSNs if all the nodes deliver their data to the sink, it results in a huge expenditure of energy. So there is a need to reduce energy consumption which enhances the performance of the network in terms of life time.

3. PROPOSED PROTOCOL

In WSN, sensor field are divided into grids in order to build diagonal path from each grid towards sink. Density of nodes is a decision factor of different energy levels of each grid. We compared our proposed protocol to the one proposed in [7]. The two protocols have some common features for example: both protocols establish multiple routing paths, they differentiate between two types of topologies (circular and grid) [8]. In the proposed WSN is divided into different grids according to information of the node location, and then the nodes are organized within the grid by the clustering way. The clustering head is chose according to energy level of the cluster nodes and going to make forcefully sleep the nodes which are below some predefined power up to all nodes in that group [9].

3.1 Energy Analysis Simulation results show that our protocol is superior in saving nodes energy and extending network lifetime. The energy saving in the proposed

protocol comes from the multiple energy aware schemes that are used. First, not all nodes participate in paths establishment phase; just one node per grid broadcasts grid based routing information whose energy level is high as compared to other nodes [10]. Secondly, detecting topology changes (such as when a grid becomes empty or when an area became congested). Third, the way in which the paths are established (diagonal for non-boundary and vertical/horizontal for boundary) aims to utilize the grids energy evenly. Finally, based on (1) paths going through densely deployed areas are preferential, this help maintaining network connectivity and extending network lifetime [11].

3.2 Multiple Routing Paths we proposed a novel way of establishing routing paths in the network. Regardless the number of nodes in the network, each master node has multiple diagonal paths towards the sink through the neighbors (depending on the neighbor grid availability that is based on the energy level of the nodes) [12].

3.3 Queue Occupancy several experiments are performed to compare the proposed protocol queue occupancy based on energy level of the nodes on grid. Energy level is a major issue due to the existence of several alternative paths [13].

Data dissemination protocol based on different groups of grid in WSN guarantees a higher level of node on grid sends data to sink by different paths going through densely deployed areas and other nodes are in sleep mode thus extending the lifetime of routing protocol in WSN.

4. PERFORMANCE

A. Energy consumption

Data dissemination protocol based on different groups of grid in WSN proposes a low energy consumption comparing to literature studies paper [14] [15]. In this each node goes to sleep mode when its energy level reach at predefined power levels and it will be in sleep mode till in a cluster each node comes on same energy level, due to this concept all the nodes are active at last & improve the life time of node in wireless sensor network.

B. Packet Reception Ratio

Data dissemination protocol based on different groups of grid in WSN presents the average packet reception ratio at sink comparing to literature studies paper. Reason is behind that all nodes are active at last and congestion not affected WSN and PRR make it average at last.

5. CONCLUSION

In this paper, data dissemination protocol based on different groups of grid in WSN is designed to address two main

important issues in wireless sensor networks: extending network lifetime and increasing the reception of packets at the sink node. Our proposed protocol extends the life time of the sensor network and utilizes the available storage. As we are planning to study the protocol performance for networks with mobile base and compare data dissemination protocol based on different groups of grid in WSN performance to other cluster based routing protocols.

6. ACKNOWLEDGEMENT

We would like to sincerely thank Professor Kanika Sharma for her sincere support and guidance in our research work. She is responsible for a great deal of whiteboard illumination and appropriate course corrections. Under her guidance we became able to achieve our target with desired performance.

REFERENCES

- [1] KazemSohraby, Danielminoli, TaiebZnati "WIRELESS SENSOR NETWORKS: Technology, Protocols, and Applications", published by John Wiley & Sons, Inc., Hoboken ew Jersey, 2007.
- [2] Al-Karaki, J.N.; Kamal, A.E. "AI-Routing techniques in wireless sensor networks: a survey" Wireless Communications, IEEE International Conference, Vol.11, No.6, pp.6 - 28, 2004.
- [3] Kazemsohraby, Daniel Minoli, Taieb Znati, "Wireless Sensor Networks", Technology, Protocols and applications, 2007.
- [4] Wu, Zhengyu; Song, Hantao; Jiang, Shaofeng; Xu, Xiaomei, "Energy-Aware Grid Multipath Routing Protocol in Mobile Ad Hoc network (MANET)" , IEEE Asia International Conference on modeling & simulation (AMS), pp. 36 - 41, 2010.
- [5] Yimei Kang; Yang Han; Jiang Hu, "A Node Scheduling Based on Partition for WSN", IEEE International Conference on Wireless Telecommunications Symposium (WTS) , pp.1 - 6, 2012.
- [6] Lambrou, T.P. Panayiotou, C.G., "A Survey on Routing Techniques Supporting Mobility in Sensor Networks", 5th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), pp. 78 - 85, 2009.
- [7] Vijayalakshmi, A.; Ranjan, P.V., "Slot Management based Energy Aware routing (SMEAR) for wireless sensor networks", IEEE International Conference on Computing, Communication and Applications (ICCCA), pp. 1 - 5, 2012.
- [8] Shanti, C.; Sahoo, A., "DGRAM: A Delay Guaranteed Routing and MAC Protocol for Wireless Sensor Networks", IEEE Transactions on Mobile Computing, Vol.9, No.10, pp.1407-1423, 2010.
- [9] Zhibin Li; Liu, P.X. "Priority-based Congestion Control in Multi-path and Multi-hop Wireless Sensor Networks", IEEE Conference on robotics & biomimetics (ROBIO), pp. 658 - 663, 2007.
- [10] Amin, A.; Mehbob, W.; Ranjha, A.H.; Abbas, H.; Abbas, N.; Anjum, W., "Efficient Load Sharing Routing Algorithm to Increase lifetime of Wireless Sensor Networks", IEEE Saudi

- International Conference on Electronics, Communications and Photonics Conference (SIEPCP), pp.1 - 5, 2011.
- [11] Sergiou, C.; Vassiliou, V., "Study of lifetime extension in wireless sensor networks through congestion control algorithms", IEEE Symposium on Computers and Communications (ISCC), pp. 283 - 286, 2011.
- [12] DijunLuo; Xiaojun Zhu; Xiaobing Wu; Guihai Chen, "Maximizing Lifetime for the Shortest Path Aggregation Tree in Wireless Sensor Networks", IEEE International Conference INFOCOM, pp.1566 – 1574, 2011.
- [13] Banimelhem, O.; Khasawneh, S., "Grid-based Multi-path with Congestion Avoidance Routing (GMCAR) Protocol for Wireless Sensor Networks", IEEE International Conference on Telecommunication (ICT) , pp. 131 - 136, 2009.
- [14] SungHwi Kim; Sang-Ha Kim, "Data Dissemination Protocol with Hole Masking Algorithm in Grid-based Wireless Sensor Networks", 4th International Conference on Ubiquitous and Future Networks (ICUFN), pp .503 - 508, 2012.
- [15] Wei-dong Liu; Zheng-dong Wang; Shen Zhang; Qing-qing Wang, "A Low Power Grid-based Cluster Routing Algorithm of Wireless Sensor Networks", IEEE International Forum on Information Technology and Applications (IFITA), Vol. 1, pp. 227 - 229, 2010.
- [16] Ayoub, Z.T.; Ouni, S.; Kamoun, F. "Energy consumption analysis to predict the lifetime of IEEE 802.15.4 wireless sensor networks", Third International Conference on Communications and Networking (Com Net), pp. 1-6, 2012.

Cluster-Based Routing Protocols for Heterogeneous Wireless Sensor Networks

Suniti Dutt¹, O. S. Khanna²

¹ME Student, ²Associate Professor

^{1,2}National Institute of Technical Teachers Training and Research, Chandigarh

¹sunitidutt@gmail.com, ²oskhanna@gmail.com

Abstract: New routing protocols are being continuously researched to eliminate the various problems currently faced by Wireless Sensor Networks. Clustering techniques aim at reducing the energy usage of each node and hence cluster based routing protocols are being explored for further improvements. At present, a major portion of this research is based on the assumption that all the sensor nodes have initially the same amount of energy. This is known as a homogeneous energy setting. In a heterogeneous environment, a certain population of the sensor nodes is furnished with additional energy resources, thus leading to an energy-hierarchy. Though the heterogeneous wireless sensor networks are more complex than the homogeneous ones, yet they achieve better performance since the use of energy-hierarchy results in a higher energy saving. In the present paper, some of the major hierarchical routing protocols for heterogeneous wireless sensor networks have been reviewed in terms of their network lifetimes.

IndexTerms: Clustering, Heterogeneous Wireless Sensor Network Hierarchical Routing.

1. INTRODUCTION

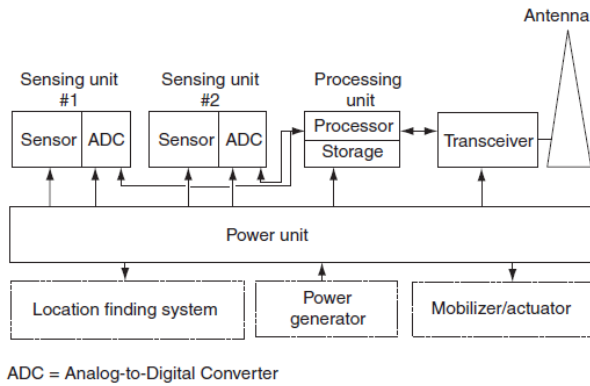


Fig. 1. A typical sensor node [1]

Wireless communication technologies are continually growing in diverse areas, which are providing new opportunities for networking and services. One such area is Wireless Sensor Networks (WSN). A Sensor Network [1] is basically comprised of sensing, computing, and communication components. This provides an administrator

with the capabilities of instrumenting, observing, and reacting to events in a given environment. The administrator can be any governmental, commercial, or industrial organization. The environment can be the physical world or any biological system. A typical sensor node consists of a sensing unit, a processing unit, a communication unit as well as a power unit as depicted in Fig. 1 [1].

There is a grave need for these sensor nodes to handle more complex functions during data aggregation and transmission. Energy saving methodologies are therefore, a major requirement for these battery-run sensor nodes. Challenges faced in WSNs due to constrained energy supply and bandwidth necessitates the need for developing energy aware protocols at all levels of protocol stack. To offer efficient energy management in WSN, various researches have focussed on sensor network hardware and energy-aware routing protocols. The various routing protocols can be broadly classified as one of either location-based, or data-centric or hierarchical routing protocols.

A. Hierarchical Routing

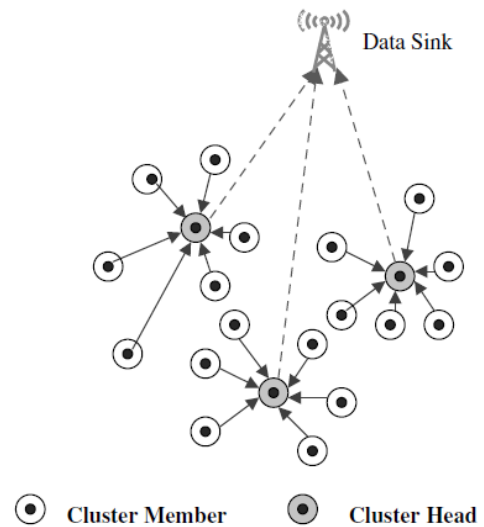


Fig. 2. A basic hierarchical architecture [1]

Hierarchical routing in WSN involves the hierarchical formation of clusters when sending information from the sensor nodes to the BS or the data sink as shown in Fig. 2 [1]. Clustering is a method by which the sensor nodes are organized hierarchically according to their relative proximity to each other. The cluster formation process is based on the residual energy of the sensor nodes and election of a Cluster Head (CH) [2].

The energy consumption in a hierarchical manner creates a very effective and reliable means of routing the sensed data from the sensor nodes to the BS. Clustering of the nodes helps to shorten the routing table, and thus, the routing is done more easily. It also helps to conserve the bandwidth because only the CHs take part in inter-cluster communication. Each sensor node first performs a routing table lookup for a CH in its region and then it routes its collected data to that CH. This CH then calculates the path estimate, which is based on the shortest distance to another CH closer to the BS or directly to the BS. In order to maintain the routing table, link information is exchanged from time to time among all the sensor nodes. Hierarchical routing thus significantly reduces the energy consumption as it employs multi-hop communication for a specific cluster. Also, it performs data aggregation and fusion in such a way that decreases the number of data units carried across the network to the sink.

B. Homogeneous v/s Heterogeneous WSNs

In [3], a comparative study of homogeneous vs. heterogeneous clustered sensor network has been carried out. The authors in [4] have described the cost benefits of deploying a heterogeneous system over a homogeneous system. A homogeneous sensor networks is defined as a network which consists of identical nodes having the same energy level, same processing capabilities, and same sensing range. On the other hand, heterogeneous sensor network consists of sensor nodes having different abilities, such as different energy level, different sensing range and different computational abilities. Most heterogeneous network may also have varying level of these mentioned abilities that depend on the deployment scenario.

There may be two major sources of energy heterogeneity; one may be due to the initial settings of the network where sensors are deployed to perform different tasks and the other may arise during the operation of the network. In many applications, there is a need for sensors with different energy levels in the network, but most likely with same processing or computational capacities. In such applications, if we have a homogeneous setting by deploying a sensor network with the nodes having same processing capabilities, computational power, and same energy level, there is tendency that some of the nodes will die out faster than the

others regardless of using the best protocol for the homogeneous setting.

This results into wastage of residual energy in the remaining alive nodes. In this case, there may be a need to introduce new sensors into the network system, which is referred to re-energizing the network system. Once the network is re-energized, it constitutes energy heterogeneity in the sensor network. Most real life cases deal with such type of energy-heterogeneity rather than homogeneity. Thus, with a proper implementation of a suitable protocol, a mixed deployment of these sensor nodes in terms of energy level can boost the performance, improve the overall network lifetime and at the same time reduce the cost by as much as possible.

C. Types of Heterogeneity

There are three types of heterogeneity that can be exploited in WSNs [5]:

- Computational heterogeneity implies that certain number of nodes in a heterogeneous WSN have more powerful processors or larger memory as compared to the rest of the nodes.
- Link heterogeneity implies that certain number of nodes in a heterogeneous WSN have a higher bandwidth and long distance network transceiver than the normal node. Link heterogeneity can provide a more reliable data transmission.
- Energy heterogeneity implies that certain number of nodes in a heterogeneous WSN have are equipped with a larger battery.

Among these three types of heterogeneity, the most significant type is the energy heterogeneity and is widely exploited to conserve the energy in a resource-restrained WSN.

2. HETEROGENEITY-AWARE ROUTING PROTOCOLS

Qing et al [6] have proposed a Distributed Energy-Efficient Clustering (DEEC) scheme based on the ideas of LEACH [7]. The theme of the protocol is to elect the CHs using probability based approach in order to estimate the ratio of the residual energy of each node and the average energy of the network. Eventually, the node with high residual energy will become CHs more often than the nodes with low energy. The objective of DEEC is to propose an energy aware routing protocol for a heterogeneous WSN. The architecture of DEEC protocol uses the initial and the residual energy of each node to select CHs. DEEC estimates the idealistic value of the network lifetime in order to avoid the global knowledge of the network. This is an advantage of DEEC. The only restraint with this scheme is that the

estimated average energy is inversely proportional on the energy consumed in each round. This proves to be a drawback in the model estimation of DEEC. The simulation results show that DEEC achieves a longer lifetime than LEACH protocol in a heterogeneous environment.

An improvement over DEEC is proposed as Stochastic DEEC (SDEEC) [8]. The CH selection is based on a node's residual energy. The Stochastic scheme reduces the intra-cluster transmission. Similar to DEEC, SDEEC considers two-level energy heterogeneity, but it conserves more energy as it puts the non-CH nodes into sleep mode. The network is divided into dynamic clusters. All non-CH sensor nodes transmit their data to their respective CHs during a given time interval. The CH always keeps its receiver in an on state so as to receive the data from its cluster members. Next, the CHs send the aggregated data of their members to the primary CH. The drawback of SDEEC is that if the non-CH sensor nodes are in the sleep mode then how are they going to know about the start of CH selection for the next round.

Smaragdakis et al [9] were one of the first to address the impact of energy heterogeneity of nodes in WSNs in the form of SEP network layer protocol. Their approach was to assign weighted probability to each node based on its energy level as the network evolves. One major characteristic of this approach is that it rotates the CH to adapt the election probability to suit the heterogeneous settings. The authors exploited the capabilities of LEACH to develop an adaptive and well distributed model to cater for extra energy introduced into the network, which is a source of heterogeneity. Under the model development of SEP, two kinds of nodes with different energy levels were used, constituting a two-level hierarchical WSN in a single-hop setting. The authors used two kinds of nodes: normal nodes and advanced nodes. The advanced nodes have more energy by a factor of α over the normal nodes. The advanced nodes take up CH position more than the normal nodes during the same epoch according to SEP model estimation. SEP used an election probability based on the initial energy of each node to elect the CHs by assigning a weight equal to the initial energy of each node divided by initial energy of the normal nodes. The weighted probabilities for normal and advanced nodes in SEP were chosen to reflect the extra energy introduced into the network system. It has been shown by simulations that SEP always extends the stability period and also increases the average throughput as compared to LEACH clustering protocol.

The authors in [10] have put forward a Distributed Energy Balance Clustering (DEBC) Protocol for heterogeneous WSNs. The CH selection is based on the ratio of the remaining energy of a sensor node and the average energy of the whole network. DEBC proposes two-level energy heterogeneity as well as multi level heterogeneity. The

simulation result show that DEBC protocol leads to a longer lifetime than existing clustering protocols.

The authors of [11] have proposed a Cluster Based Service Discovery (CBSD) protocol for heterogeneous WSNs. The aim of CBSD is to design a service discovery protocol which cuts down the workload of the resource-restrained sensor nodes in a heterogeneous WSN. The authors propose the formation of clusters based on the capabilities of sensor nodes. In this approach, each sensor node is allotted a unique hardware identifier and a weight, which is the grade of its capability. Nodes with higher capability grades are more likely to become the CH. These nodes then behave as a distributed directory of service registrations for all the cluster members. Thus, communication expenses are lowered since the service discovery messages are interchanged only among the directory nodes. The CBSD algorithm reacts very quickly to any topological changes in the WSN as it makes the decisions based on single-hop neighbourhood information.

Another SEP protocol has been put forth by Bala et al [12], known as Deterministic-SEP (D-SEP), for electing CHs in a distributed manner in two-, three-, and multi-level hierarchical WSNs. The authors have elucidated the CH selection procedure by detailing the threshold and probability equations. D-SEP protocol accomplishes to enhance the lifetime and stability of the WSN in a heterogeneous environment. Since CHs consume more energy than cluster members, the role of CH is rotated among sensor nodes, as in other protocols. At each round, a node decides whether to become a CH or not, which is based on threshold computed by the specified percentage of CHs for the network and the number of times that node has been a CH so far. This decision is made by the sensor nodes themselves by choosing a random number lying between 0 and 1. If this chosen random number is less than a particular threshold, this sensor node becomes a CH for the current round. In order to reach a constructive conclusion, different cases of two-level and three-level heterogeneous environments have been reported and compared with SEP. D-SEP protocol goal is to increase the lifetime and stability of the network in the presence of heterogeneous nodes. Since CHs consume more energy than cluster members in receiving and sensing data from their member nodes, performing signal processing and sending the aggregated data to next node or BS, the role of CH must be rotated among sensor nodes. Therefore, D-SEP works in rounds as SEP and also considers how to optimally select the CHs in the heterogonous network. Traditionally as per SEP, CH algorithm is broken into rounds. At each round node decides whether to become a CH based on threshold calculated by the suggested percentage of CHs for the network (determined a priori) and the number of times the node has been a CH so far. This decision is made by the nodes by choosing the random number between 0 and 1. If the number

is less than a threshold $T(s_i)$ the node becomes a CH for the current round. In the proposed D-SEP the threshold is modified. In a Two-level Heterogeneity, two type of nodes known as normal and advanced nodes are considered with their different initial energy for two-level heterogeneous networks. The reference value of p_i is different for these types of nodes. In case of Three-level Heterogeneity, three types of nodes known as normal, advanced and super nodes are considered based on fractional difference in their initial energy level. In the multi-level heterogeneity all the nodes have been considered with different initial energy. Significant improvement has been shown using D-SEP over SEP in terms of network lifetime, energy consumption and data transmission to BS.

In [13], Zhou et al proposed a Stable Election Protocol based on Energy Dissipation Forecast Method (EDFM) for Clustered HWSNs. Simulation results show that EDFM equilibrates the energy expenditure better than the established routing protocols, and also extend the lifetime of HWSNs. The same authors propose another framework with energy and computational heterogeneity for HWSNs in [14]. The energy dissipation model and the optimum number of clusters in HWSNs are also demonstrated with a mathematical model, providing the direction for designing the clustering protocols. Furthermore, a novel energy-efficient algorithm that guarantees authentic and reliable transmission for HWSNs is projected, to improve the clustering strategy in LEACH and LEACH-like protocols, in which the CH selection procedure is based on a method of Energy Dissipation Forecast and Clustering Management (EDFCM). EDFCM considers the remainder energy and energy expenditure rate in all the nodes. Matlab Simulations show that EDFCM balances the power consumption much better than in the conventional routing algorithms and also prolongs the network lifetime.

Parvin has proposed a Hierarchical Energy Aware Routing Protocol (HEARP) [15], which is based on LEACH and PEGASIS routing protocols. HEARP protocol is divided into rounds, with each round consisting of a setup phase and a data transmission phase. Clusters are formed and CHs are chosen, similar to LEACH. A chain is then established among the chosen CHs, similar to PEGASIS protocol. A leader is then chosen from amongst the chain members to transmit data to the BS. In [16], the authors have put forward a Weighted Election Protocol (WEP) in order to increase the stability region. It is a combination of the SEP and HEARP protocols. The Simulation results show that WEP prolongs the stability period as compared to LEACH, HEARP and SEP protocols.

Base Station Initiated Dynamic Routing Protocol was proposed by Verma et al in [17]. In this scheme, the nodes that are stronger as compared to other nodes of the WSN in terms of energy, computational capacity and location

awareness, act as the CHs. The transmission powers of the CHs are adjusted so that only single hop communication is possible. A level is defined in this scheme which designates the distance of a CH from the BS. A lower level stands for a CH that is closer to the BS and if the level is high, it signifies that the CH is away from BS. Information flow always occurs from a higher level to a lower level. The BS sets its level as zero and then broadcasts a data packet for initiation to the CHs. Since the CHs are at different signal strength as compared to the normal nodes, they receive the data packet and set their levels according to it. When all the CHs of first level have been selected, they broadcast a packet telling others about their level. The CHs at lower levels choose their parent from among the upper level CHs. This procedure is repeated till all CHs get connected. The CHs now broadcast a message that all normal sensor nodes should then join the CH according to the RSS (Radio Signal Strength).

A Distributed Election Clustering Protocol (DECP) for Heterogeneous WSNs has been proposed in [18] by X. Wang et al. DECP selects the CH based on the remaining energy of the nodes as well as the communication cost. DECP is based on Average Power Distinction (APD) to assess the power levels of the sensor nodes. Thus, the nodes with a higher residual energy have more prospects of becoming a CH as compared to nodes with less energy. All the nodes broadcast their current energies as well as collect the energy information messages sent by the other nodes. When all the nodes have required energy and distance information about their neighbours, they calculate the communication costs and broadcast these to their neighbours. The CHs are then selected based on the minimum costs. Simulation results show that DECP provides an improved stability region as compared to other cluster based protocols.

3. CONCLUSIONS AND FUTURE DIRECTIONS

It is observed that one of the main challenges faced by the resource-constrained WSNs is to increase the network lifetime. Clustering is a cardinal topology control mechanism, based on which we can handle the challenges posed by WSNs. The present protocols majorly assume homogeneous energy utilisation of each node with respect to the total energy of the network. However, these protocols are not heterogeneity-aware, in the sense that whenever there is an energy difference between nodes in the network, the sensor nodes die out much more rapidly. In real life situations, it is difficult for the sensor nodes to maintain their energy uniformly, resulting in an energy imbalance between them. Other clustering schemes were thus explored, which assume that some percentage of the population of sensors is furnished with extra energy resources—this is the origin of heterogeneity. Some of the major routing protocols for a heterogeneous energy setting have been reviewed in this

paper, specifically with respect to their network lifetimes, stating their strengths and limitations. However, there is still much work to be done other than improvement in the network lifetimes. Further research is required to address issues like the Quality of Service (QoS) of the WSN, deployment of mobile sensor nodes and secure data transmission process among various others issues.

REFERENCES

- [1] K. Sohraby, D. Minoli, T. Znati (2007). *Wireless Sensor Networks: Technology, Protocols, and Applications*, John Wiley & Sons, pp 1-38.
- [2] Kemal Akkaya and Mohammed Younis, "A survey on routing protocols for wireless sensor networks and Ad hoc networks", *Adhoc Networks Journal Elsevier*, vol. 3 pp325 -349, 2005.
- [3] VivekMhatre and Catherine Rosenberg, "Homogeneous vs. Heterogeneous Clustered Sensor Networks: A comparative study", *IEEE International Conference on Communications (ICC)*, pp 3646-3651, 2004.
- [4] Chun-Hsien Wu and Yeh-Ching Chung, "Heterogeneous Wireless Sensor Network Deployment and Topology Control Based on Irregular Sensor Model" 2nd International Conference on Advances in Grid and Pervasive Computing (GPC), Springer-Verlag, pp. 78-88, 2007.
- [5] V. Katiyar, N. Chand and S. Soni, "Energy-Efficient Multilevel Clustering in Heterogeneous Wireless Sensor Networks", *International Conference on Advances in Computing, Communication and Control*, Springer, pp 293-299, 2011.
- [6] Li Qing, Qinxin Zhu, and Mingwen Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks", *Computer Communications Journal Elsevier*, vol. 29, issue 12, pp 2230-2237, 2006.
- [7] Wendi R. Heinzelman, AnanthaChandrasekaran, and HariBalakrishnan, "Energy efficient communication protocol for wireless microsensor networks", *IEEE International Conference on System Sciences*, pp 1-10, 2000.
- [8] B. Elbhiri, R. Saadane, D. Aboutajdine, "Stochastic Distributed Energy-Efficient Clustering (SDEEC) for Heterogeneous Wireless Sensor Networks", *ICGST International Journal on Computer Network and Internet Research, CNIR*, vol. 9, issue 2, pp 11-17, 2009.
- [9] GeorgiosSmaragdakis, Ibrahim Matta, and AzerBestavros, "SEP: A Stable Election Protocol for clustered heterogeneous wireless sensor networks", *International Workshop on Sensor and Actor Network Protocols and Applications (SANPA)*, 2004.
- [10] C. Duan, H. Fan, "Distributed Energy Balance Clustering (DEBC) Protocol for HWSNs", *International Conference on Wireless Communications, Networking and Mobile Computing (WICOM)*, pp 2469-2473, 2007.
- [11] Marin Perianu, J. Scholten, P. Havinga, P. Hartel, "Cluster-Based Service Discovery for Heterogeneous Wireless Sensor Networks", *International Journal of Parallel, Emergent and Distributed Systems*, Vol. ,Issue. , 1-35, 2007.
- [12] ManjuBala and LalitAwasthi, "Proficient D-SEP Protocol with Heterogeneity for Maximizing the Lifetime of Wireless Sensor Networks", *International Journal of Intelligent Systems and Applications*, vol 7, pp 1-15, 2012.
- [13] Haibo Zhou, Yuanming Wu, and GuangzhongXie, "EDFM : A Stable Election Protocol based on Energy Dissipation Forecast Method for Clustered Heterogeneous Wireless Sensor Networks", *5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp 1-4, 2009.
- [14] Haibo Zhou, Yuanming Wu, Yanqi Hu and GuangzhongXie, "A novel stable selection and reliable transmission protocol for clustered heterogeneous wireless sensor networks", *Computer Communications Journal Elsevier*, vol. 33, pp 1843-1849, 2010.
- [15] S. Parvin, "Hierarchical Energy Aware Routing Protocol (HEARP) for Wireless Sensor Networks", *M.Sc. Thesis*, University of Rajshahi, Bangladesh.
- [16] Md. G. Rashed and M. H. Kabir, "Weighted Election Protocol for Clustered Heterogeneous Wireless Sensor Networks", *Journal of Mobile Communication*, Vol. 4, Issue 2, pp 38-42, 2010.
- [17] S. Varma, N. Nigam and U. Tiwary, "Base Station Initiated Dynamic Routing Protocol", *IEEE International Conference of Wireless Communications and Sensor Networks*, pp 1-6, 2008.
- [18] X. Wang and G. Zhang, "DECP: A Distributed Election Clustering Protocol for Heterogeneous Wireless Sensor Networks", *Springer International Conference on Computational Science (ICCS)*, pp 105-108, 2007.

Security Issues in Ad Hoc Networks: A Survey

Rumisa Firdous¹, Emmanuel S. Pilli², Shabir Ahmad Sofi³

^{1,2}Graphic Era University, Dehradun, India,

³NIT Srinagar, Kashmir, India

¹rumisafirdous@gmail.com, ²emmshub@gmail.com, ³shabir@nitsri.net

Abstract: Black hole attack is one of the most serious security problems in MANET. In this paper we introduced some of the proposed works in detecting black hole attacks. We compared these methods and observed that most of these algorithms suffer from overload and low speed which is a research area for finding efficient solutions against these attacks.

Index Terms: MANETs, Security, Black hole attack.

1. INTRODUCTION

Wireless networks can be either infrastructure based networks or infrastructure less networks. The infrastructure based networks uses fixed base stations, which are responsible for coordinating communication between the mobile hosts (nodes). The ad hoc networks falls under the class of infrastructure less networks, where the mobile nodes communicate with each other without any fixed infrastructure between them. An ad hoc network is a collection of nodes that do not rely on a predefined infrastructure to keep the network connected. An ad-hoc network can change its form depending on the work on hand.

A MANET is an infrastructure-less network consisting of set of mobile nodes or mobile devices wishing to communicate with each other via shared wireless medium; it does not have any centralized administration and therefore, line of defense is pretty unclear. Each node has limited communication range in the network and it node acts as a router to forward packets to another node. It is rapidly deployable and highly adaptive in nature. Nodes have high mobility and communication is done via radio broadcast medium.

Therefore, MANETs are widely used in applications such as military communication by soldiers, automated battlefields, emergency management teams to rescue, search by police or fire fighters, replacement of fixed infrastructure in case of earthquake, floods, fire etc., quicker access to patient's data from hospital database about record, status, diagnosis during emergency situations, remote sensors for weather, voting systems, sports stadiums, mobile offices, vehicular computing, electronic payments from anywhere, education systems, conference meetings, peer to peer file sharing systems [1].

2. FEATURES OF AD HOC NETWORK

The wireless ad hoc network has the following typical features [2]:

- Unreliability of wireless links between nodes. Because of the limited energy supply for wireless nodes and the mobility of the nodes, the wireless links between mobile nodes in the ad hoc network are not consistent for the communication participants.
- Constantly changing topology. Due to the continuous motion of nodes, the topology of the mobile ad hoc network changes constantly: the nodes can continuously move into and out of the radio range of the other nodes in the ad hoc network, and the routing information will be changing all the time because of the movement of the nodes.
- Lack of incorporation of security features in statically configured wireless routing protocol not meant for ad hoc environments. Because the topology of the ad hoc networks is changing constantly, it is necessary for each pair of adjacent nodes to incorporate in the routing issue so as to prevent some kind of potential that try to make use of vulnerabilities in the statically configured routing protocol.

3. SECURITY IN AD HOC NETWORKS

Mobile Ad hoc Network is constructed from a collection of nodes that can move anywhere and anytime in different areas without any infrastructure. Due to wireless communication, dynamic topology, limited resources and lack of centralized administration, MANETs are vulnerable to various types of attacks.

A. Denial of Service (Dos) attacks

DoS attacks are active attacks in which malicious nodes generate false messages in order to disrupt the network's operations or to consume other nodes' resources. DoS attack aims to crab the availability of certain node or even the services of the entire ad hoc networks. In the traditional wired network, the DoS attacks are carried out by flooding some kind of network traffic to the target so as to exhaust the processing power of the target and make the services

provided by the target become unavailable. Nevertheless, it becomes not practical to perform the traditional DoS attacks in the mobile ad hoc networks because of the distributed nature of the services. Moreover, the mobile ad hoc networks are more vulnerable than the wired networks because of the interference-prone radio channel and the limited battery power. In the practice, the attackers exactly use the radio jamming and battery exhaustion methods to conduct DoS attacks to the mobile ad hoc networks, which well correspond to the two vulnerabilities [3].

B. Impersonation

As there is no authentication of data packets in current ad hoc network, a malicious node can launch many attacks in a network by masquerading as another node i.e. spoofing. Spoofing is occurred when a malicious node misrepresents its identity in the network (such as altering its MAC or IP address in outgoing packets) and alters the target of the network topology that a benign node can gather. As for example, a spoofing attack allows forming loops in routing packets which may also result in partitioning network [4].

C. Eavesdropping

Eavesdropping is another kind of attack that usually happens in the mobile ad hoc networks. The goal of eavesdropping is to obtain some confidential information that should be kept secret during the communication. The confidential information may include the location, public key, private key or even passwords of the nodes. Because such data are very important to the security state of the nodes, they should be kept away from the unauthorized access [3].

D. Black hole attack

A basic Byzantine attack is a black hole attack where the adversary stops forwarding data packets, but still participates in the routing protocol correctly.

E. Wormhole attack

The wormhole attack is a kind of tunneling attack which is extremely dangerous and damaging to defend against. This attack occurs when two adversaries cooperate to tunnel packets between each other in order to create a shortcut (or wormhole) in the network

F. Modification attack

Modification is a type of attack when an unauthorized party not only gains access to but tampers with an asset. For example a malicious node can redirect the network traffic and conduct DoS attacks by modifying message fields or by forwarding routing message with false values [4].

G. Fabrication attack

Fabrication is an attack in which an unauthorized party not only gains the access but also inserts counterfeit objects into

the system. In MANET, fabrication is used to refer the attacks performed by generating false routing messages. Such kind of attacks can be difficult to verify as they come as valid constructs, especially in the case of fabricated error messages that claim a neighbor cannot be contacted [4].

H. Attacks against Routing

Routing is one of the most important services in the network; therefore it is also one of the main targets to which attackers conduct their malicious behaviors. In the mobile ad hoc networks, attacks against routing are generally classified into two categories: attacks on routing protocols and attacks on packet forwarding/delivery [5]. Attacks on routing protocols aim to block the propagation of the routing information to the victim even if there are some routes from the victim to other destinations. Attacks on packet forwarding try to disturb the packet delivery along a predefined path. The main influences brought by the attacks against routing protocols include network partition, routing loop, resource deprivation and route hijack [5].

4. BLACK HOLE ATTACK

A Black Hole attack is a kind of denial of service where a malicious node can attract all packets by falsely claiming a fresh route to the destination and then absorb them without forwarding them to the destination. If the malicious node replies to the requesting node before the genuine node replies, a false route will be created. Therefore, packets do not reach to the specified destination node; instead, the malicious node intercepts the packets, drops them and thus, network traffic is absorbed [6] [8].

5. DETECTION/PREVENTION OF BLACK HOLE ATTACK

A. Multipath Routing Technique

Many sensor network routing protocols are quite simple, and for this reason are sometimes even more susceptible to attacks against general ad hoc routing protocols. One such type of attack is selective forwarding attack. In a selective forwarding attack, malicious nodes may refuse to forward certain messages and simply drop them, ensuring that they are not propagated any further. A simple form of this attack is when a malicious node behaves like a black hole and refuses to forward every packet it sees. However, such an attacker runs the risk that neighboring nodes will conclude that she has failed and decided to seek another route. Karlof et al [7] proposed multipath routing can be used to counter selective forwarding attacks. Messages routed over n paths whose nodes are completely disjoint are completely protected against selective forwarding attacks involving at most n compromised nodes and still offer some probabilistic protection when over n nodes are compromised. However complete disjoint paths may be difficult to create. The

algorithm fails to suggest a method to isolate the attacking node and remove from the network.

B. Reply packet authenticity approach

Muhammad Al-Shurman et al [8] proposed two different approaches to solve the black hole attack. The first solution the sender node needs to verify the authenticity of the node that initiates the RREP packet by utilizing the redundancy of the network. The idea of this solution is to find more than one route for the destination. The SN unicast the ping packet using different routes. The IN or destination node or malicious node will ping requests. The second solution is to store the last sent packet sequence number and the last received packet sequence number in the table. It is updated when any packet is arrived or transmitted. When node receives reply from another node it checks the last sent and received sequence number. If there is any mismatch then an ALARM indicates the existence of a black hole node. This method can cause routing delay, since a node has to wait for a RREP packet to arrive from more than two nodes.

C. Anomaly detection

Satoshi Kurusawa et al. [9] used an anomaly detection scheme. It uses dynamic training method in which the training data is updated at regular intervals. Multidimensional feature vector is defined to express state of the network at each node. Each dimension is counted on every time slot. It uses destination sequence number to detect attack. The feature vector include number of sent out RREQ messages, no of received RREP messages, the average of difference of destination sequence number in each time slot between sequence number of RREP message and the one held in the list. They calculate mean vector by calculating some mathematical calculation. They compare difference between the mean vector and input data sample. If distance is greater than some threshold value then there is an attack. The updated data set to be used for next detection. Repeating this for time interval T anomaly detection is performed. This method may make mistake when a node is not malicious, but according to its higher sequence number may entered into blocked list.

D. Distributed and cooperative procedure to detect black hole attack

Chang Wu Yu et al [10] proposed a distributed and cooperative procedure to detect black hole node. In this each node detects local anomalies. It collects information to construct an estimation table which is maintained by each node containing information regarding nodes within power range. This scheme is initiated by the initial detection node which is first broadcast and then it notifies all one-hop neighbors of the possible suspicious node. They cooperatively decide that the node is suspicious node. Immediately after the conformation of black hole, the global

reaction is activated to establish proper notification system to send warning to the whole network. The simulation results show the higher black hole detection rate and achieves better packet delivery, when the network is busier it achieves less overhead.

E. Data Routing Information (DRI) and cross checking technique

S. Ramaswamy [11] et al presented an algorithm which claims to prevent the cooperative black hole attacks in ad hoc network. In this algorithm each node maintains an additional Data Routing Information (DRI) table. Whenever a node (say IN) responded to a RREQ it sends the id of its next hop neighbor (NHN) and DRI entry for NHN to the source. If IN is not a trustable node for source then source sends a further route request (FRq) to NHN. NHN in turn responds with FRp message including DRI entry for IN, the next hop node of current NHN, and the DRI entry for the current NHN's next hop. If NHN is trusted node then source checks whether IN is a black hole or not using the DRI entry for IN replied by NHN. If NHN is not trustable node then the same cross checking will be continued with the next hop node of NHN. This cross checking loop will be continued until a trusted node is found. Moreover, in the case when the network is not under the attack, the algorithm takes more time to complete. If there is not any attack in the network, this scheme works very slowly and has a huge overhead for checking all nodes in a route.

F. Watchdog/Pathrater technique

Marti.et.al [12] proposed a technique to trace malicious nodes by using watchdog/pathrater. In watchdog when a node forwards a packet, the node's watchdog verifies that the next node in the path also forwards the packet by promiscuously listening to the next node's transmissions. If the watchdog finds the next node does not forward the packet during a predefined threshold time, the watchdog will accuse the next node as a malicious node to the source node. The proposal has two short comings:

- 1) to monitor the behavior of nodes two or more hops away, one node has to trust the information from other nodes, which introduces the vulnerability that good nodes may be bypassed by malicious accusation; 2) The watchdog cannot differentiate the misbehavior from the ambiguous collisions, receiver collisions, controlled transmission power, collusion, false misbehavior and partial dropping. In pathrater algorithm each node uses the watchdog's monitored results to rate its one-hop neighbors. Further the nodes exchange their ratings, so that the pathrater can rate the paths and choose a path with highest rating for routing. The algorithm fails to detect the attacker in the presence of selective forwarding attack and to completely remove the node from the network.

G. Extension of Watchdog technique

This method is an extension similar to watchdog design [13]. It categorizes nodes into two groups called trusted and ordinary. Trusted nodes previously proved their trustfulness to other nodes. Watchdog nodes that monitor the network are selected from these trusted nodes.

Watchdog nodes are selected according to some criterion such as energy of each node, enough storage memory, and node calculating power. Watchdog tasks exchange between trusted nodes after a period of time. In each watchdog two limit values and counters are considered, Acceptance threshold and Suspect threshold. Acceptance threshold is a limit that once correct packet sending of one node exceeds it, that node enters in trusted nodes. Suspect threshold is used to count maliciousness of one node for packet dropping and after exceeding that limit, that node enters in malicious nodes and announces that as a black hole node to the network.

H. Local collaboration and cross validation technique

SCAN [14] exploits two ideas to protect the mobile ad hoc networks

Local collaboration: The nodes monitor each other and also sustain routing tables of each other. Each node uses token that authenticates itself to the network. If one node is suspected to be malicious, other nodes revoke its token and alert token revocation to all nodes in the network and they insert that node in their token revocation list. So, the malicious node does not have any access to the network.

Information cross-validation: Each node monitors its neighbors by cross-checking the overheard transmissions, and the monitoring results from different nodes are further cross validated. As a result, the security solution is self-organized, distributed, and fully localized. In SCAN once a malicious node is convicted by its neighbors, the network reacts by depriving its right to access the network by revoking its token. Due to mobility of nodes, routing tables change and mistakes in finding malicious nodes will be increased. Also this method needs renewal of table entry of neighbors in certain period of time.

I. Information and cross checking using *FREQ* and *FREP*

Hesiri Weerasinghe [15] proposed the solution which discovers the secure route between source and destination by identifying and isolating cooperative black hole nodes. This solution adds on some changes in the solution proposed by the Ramaswamy to improve the accuracy. This algorithm uses a methodology to identify multiple black hole nodes working collaboratively as a group to initiate cooperative black hole attacks. This protocol is a slightly modified version of AODV protocol by introducing Data Routing

Information (DRI) table and cross checking using Further Request (FREQ) and Further Reply (FREP). The simulation result shows that the AODV highly suffer from cooperative black hole in terms of throughput and packet losses. The performance of the solution is good and having better throughput and minimum packet loss percentage over other solutions.

J. Counter Black hole attack against the AODV routing protocol

N.H. Mistry in [16] has proposed for the source node to verify the RREP destination sequence number by analyzing the RREP messages which arrived within the predefined waiting period by using the heuristic method. If the sequence number is found to be exceptionally high, the sender of the respective RREP will be marked as malicious node. The major issue in this method is the latency time during the route discovery process since the source node has to wait until the waiting time period expired before the routing table can be updated. In the event where there is no attack in the network, the node still suffers with the latency time.

6. CONCLUSION

We conclude that in black hole attack a malicious node advertises that it has a fresher route to the destination and sends reply to the source node before other nodes send a reply. Thus attacker node attracts all traffic in its transmission range towards itself and drops packets causing packet loss. We discussed various approaches as solution to black hole attack. Of all the solutions we analyze that that Information DRI and cross checking using FREQ and FREP approach using a methodology to identify multiple black hole nodes working collaboratively to initiate cooperative black hole attacks is having better throughput and minimum packet loss over other solutions.

REFERENCES

- [1] Nadia Qasim, Fatin Said, and Hamid Aghvami, "PerformanceEvaluation of Mobile Ad Hoc Networking Protocols", World Congress on Engineering, 2008, pp. 219-229
- [2] Amitabh Mishra and Ketan M. Nadkarni, Security in Wireless Ad Hoc Networks, in Book TheHandbook of Ad Hoc Wireless Networks (Chapter 30), CRC Press LLC, 2003.
- [3] Wenjia Li and Anupam Joshi Security Issues in Mobile Ad Hoc Networks - A Survey.
- [4] Tanu Preet Singh, Satinder Kaur, Security Threats in Mobile Adhoc Network: A Review, IRACST International Journal of Computer Networks and Wireless Communications (IJCNWC), ISSN: 2250-3501 Vol. 2, No. 1, 2012
- [5] Yongguang Zhang and Wenke Lee, Security in Mobile Ad-Hoc Networks, in Book Ad Hoc Networks Technologies and Protocols (Chapter 9), Springer, 2005.

-
- [6] Anu Bala, Munish Bansal and Jagpreet Singh, "Performance Analysis of MANET under Blackhole Attack", First International Conference on Networks & Communications, 2009, pp. 141-145.
 - [7] Chris Karlof, David Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures", 2003
 - [8] Mohammad Al-Shurman, Seong-Moo Yoo and Seungjin Park, "Black Hole Attack in Mobile Ad Hoc Networks", ACMSE, April 2004, pp.96-97.
 - [9] Satoshi Kurosawa, Hidehisa Nakayama, Nei Kato¹, Abbas Jamalipour, and Yoshiaki Nemoto¹, "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method", International Journal of Network Security, vol.5 no.3, Nov. 2007, pp.338-346.
 - [10] Chang Wu Yu, Tung-Kuang Wu, Rei Heng Cheng, and Shun Chao Chang, "A Distributed and Cooperative Black Hole Node Detection and Elimination Mechanism for Ad Hoc Network", Springer-Verlag Berlin Heidelberg, 2007.
 - [11] Sanjay Ramaswamy, Huirong Fu, Manohar Sreekantaradhya, John Dixon and Kendall Nygard, "Prevention of Cooperative Black Hole Attack in Wireless Ad Hoc Networks".
 - [12] S. Marti, T.J. Giuli, K. Lai, M. Baker, Routing Misbehavior in Mobile Ad Hoc Networks. Proceedings of the 6th annual international conference in Mobile Computing and Networking (MOBICOM), Boston, Massachusetts, United states, 2000.
 - [13] Patcha, A., and Mishra, A., 2003. Collaborative security architecture for black hole attack prevention in mobile ad hoc networks. In Proceedings of the Radio and Wireless Conference (RWCON), VA, USA, 75-78.
 - [14] Yang, H., Shu, J., Meng, X., and Lu, S. 2006. SCAN: Self-organized network-layer security in mobile ad hoc networks, J. IEEE Selected Areas in Comm. Vol. 24, No.2 (Feb. 2006), 261-273.
 - [15] Hesiri Weerasinghe, "Preventing Cooperative Black Hole Attacks in Mobile Ad Hoc Networks: Simulation Implementation and Evaluation", International Journal of Software Engineering and Its Applications, Vol.2, No. 3, July, 2008, pp: 39-54
 - [16] Nital Mistry, Devesh C. Jinwala and Mukesh Zaveri, "Improving AODV Protocol against Black hole Attacks", International Multiconference of Engineers and Computer Scientists 2010, vol. 2, March 2010.

Review of Chain Based Hierarchical Protocols in Wireless Sensor Network

Richa Mehta¹, Sandeep Verma², O.S. Khanna³

^{1,2,3}ECE Department, NITTTR, Chandigarh, Mohali, India

¹rmrichamehta7@gmail.com, ²sandi287@gmail.com, ³oskhanna@gmail.com

Abstract: Wireless Sensor Network consists of large number of wireless sensors that are effective for collecting or gathering the data in variety of environments. It is a challenging task for designing an efficient routing scheme that can offer long network lifetime, high energy efficiency and can minimize the delay. Since the sensor nodes operates on battery of limited power. In this paper, we study various routing protocols that form the chain. Various chain based protocols such as PEGASIS, PDCH, EAPHRN, CHIRON and CRBCC are analyzed.

Keywords: WSN; chain; routing protocols; network lifetime; energy efficiency.

1. INTRODUCTION

Advanced research in wireless communications have developed the multifunctional, low cost, low power, sensor nodes which are small in size and these sensor nodes communicate in short distances. In recent years, Wireless sensor networks have achieved a great attention. A WSN consists of battery powered and resource constraint sensor nodes those are randomly deployed for sensing and collecting the data from the nearby surroundings and further reporting the information to the remote base station. Wireless sensor nodes consist of sensing, processing and communicating components.

Large numbers of wireless sensor nodes are deployed densely in WSN. An important feature of sensor network is that instead of sending raw data to the nodes, they transmit only partially processed and required data by using their processing capabilities. Wireless sensor networks are employed in some of the applications areas such as health, military, habitat monitoring and disaster supervising.

This paper is organized as follows: Section 2 discusses the architecture of wireless sensor network. In section 3, the design factors of WSN are discussed. In section 4 various applications of wireless sensor network are studied. Section 5 discusses the chain formation in WSN. In Section 6 chain based routing protocols are analyzed. The conclusion remarks follows in Section 7.

2. ARCHITECTURE OF WSN

Wireless sensor nodes are scattered in a sensor field and they have the capability to collect the data and then route the data back to the sink by using multi hop infrastructure-less architecture as shown in Fig. 1. Then, further the sink communicates with the task manager node or user through Internet. Design of wireless sensor network depends on some factors such as transmission media, scalability, production cost, operating environment, fault tolerance, hardware constraints and the power consumption [1].

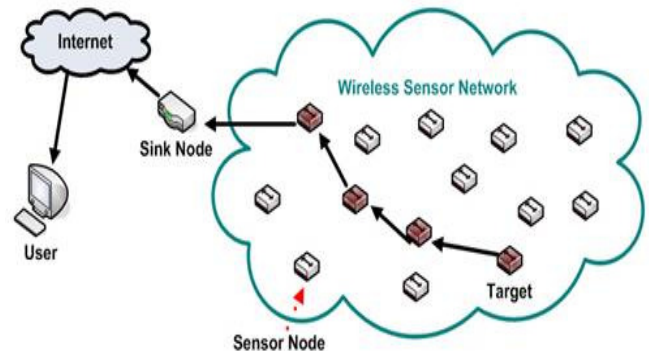


Fig. 1. Sensor nodes scattered in a sensor field [1]

3. PROPERTIES OF SENSOR NETWORK PROTOCOL

For designing good routing protocols for WSN, it is quite important to study the parameters which are relevant for the sensor applications [2].

There are various ways by which the properties of network routing protocol can be evaluated and they are:

I. Ease of Deployment

WSN contains thousands of sensor nodes that are required to be deployed in dangerous or remote environments for extracting the information which would have been otherwise impossible. Three different phases of deployment of wireless sensor nodes are:-

- *Phase-I: Pre deployment and deployment phase*

Deployment of sensor nodes can be done in a mass or one by one. If the sensor nodes are to be placed one by one then it is usually done by human or robot. If the sensor nodes are to be deployed in a mass then it can be done by dropping from rocket or plane.

- *Phase-II: Post-deployment phase*

If the sensor nodes have been deployed, then any changes in the position of sensor nodes, its reliability, and energy will change the topology.

- *Phase-III: Redeployment phase*

According to the requirement, redeployment of sensor nodes can be done [1].

II. System Lifetime

Since it is impossible to recharge the sensor node batteries therefore the WSNs must prolong for as long as possible. In order to enhance the network lifetime, protocols should be designed in such a way so that they are energy efficient [2].

III. Latency

Since the information obtained from the wireless sensor networks is time sensitive, therefore it is necessary to receive the information in a timely manner [2].

IV. Quality

Quality of WSN depends on the quality of the aggregated information set; therefore, the protocols should be designed in order to optimize a unique and application specific quality of WSN [2].

4. APPLICATIONS OF WSN

I. Industrial Applications

Industrial applications of Wireless Sensor Network as shown in Fig.2 provides:-

- Public safety
- Improves the preventive maintenance programs
- Helps to automate the data acquisition from the remote sensors
- Conservation
- Efficiency
- Control [3]

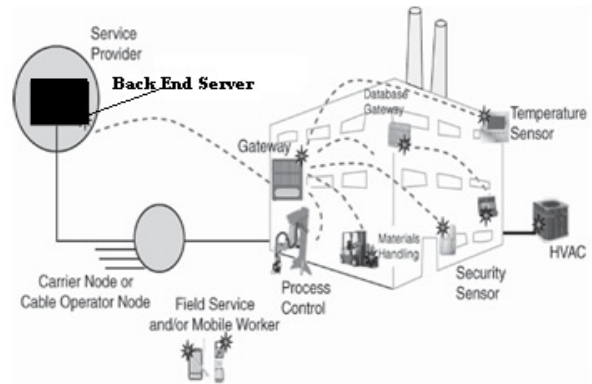


Fig. 2. Industrial control applications [3]

II. Home Applications

Home applications of Wireless Sensor Network provides:-

- Easy management of heating and cooling systems in home. Safety, conservation and control.
- Can capture water and gas utility usage data.
- Sensing applications in order to optimize the natural resources consumption.
- Installation and up gradation of home control system [3].

III. Military Applications

Various military applications in wireless sensor network are monitoring friendly resources, targeting, nuclear, biological and chemical attack [3].

IV. Medical Applications

Tracking as well as monitoring the patients and doctors, drug administration [3] as shown in Fig.3.

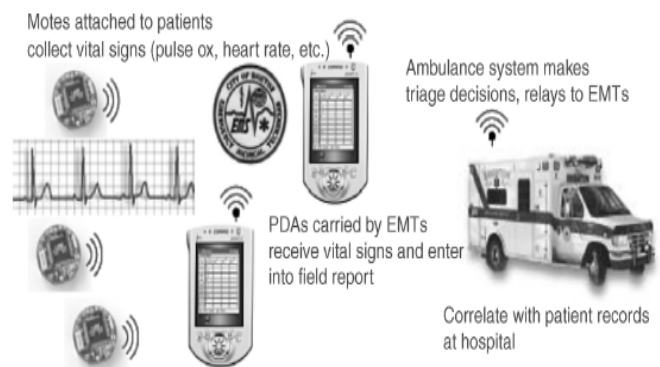


Fig. 3. Use of Code Blue for emergency response: PDA displaying real time vital signs of multiple patients [3]

5. CHAIN FORMATION IN WSN

In WSN, chain is formed by connecting different wireless sensor nodes as shown in Fig.4. One of the important applications of WSN is data collection and an efficient way for collection of data is chaining. PEGASIS i.e. Power Efficient Gathering in Sensor Information Systems was the first protocol that was devised for the concept of chaining.

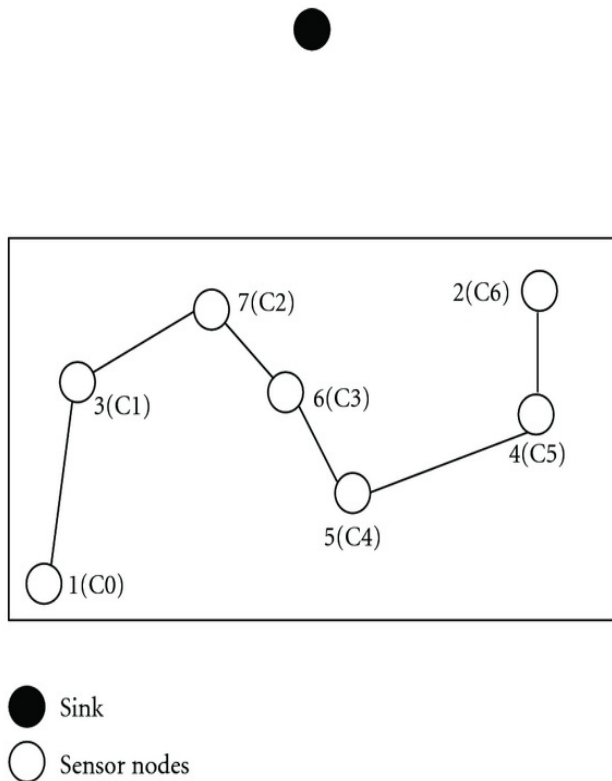


Fig. 4. Chain formation in Wireless Sensor Network [4]

Chain based routing is a significant routing scheme in which the sensor nodes are linked in advance into single or multiple chains. Each node in data dissemination phase communicates with closest neighbors and then takes the turns for transmitting the aggregated data to the base station. Even though the chain based routing schemes balances the node energy effectively and enhances the network lifetime but causes the redundant paths and transmission delays [4].

6. CHAIN BASED ROUTING PROTOCOLS

According to the current situation of wireless sensor network, many excellent and effective routing algorithms have been proposed. Some of them adopt cluster based system and the basic cluster based protocol is LEACH [5-7] and some of them adopt chain based system and the basic chain based protocol is PEGASIS [8-9].

A. Power Efficient Gathering in Sensor Information System (PEGASIS)

The concept PEGASIS algorithm is based on LEACH. The main idea in PEGASIS is formation of chain among all the sensor nodes so that each node can transmit to and receive from the closest neighbor. Aggregated data moves from one node to another node, then gets fused and finally chain leader transmits to the base station. Sensor nodes take turns while transmitting to the base station so that there is reduction in average energy spent per round by each node. PEGASIS uses greedy algorithm for the construction of chain.

Some of the advantages of PEGASIS algorithm are:-

- Sensor nodes communicate only with the neighboring nodes.
- Distance between the connected nodes has been reduced.
- Some of the disadvantages of PEGASIS algorithm are:-
- Time delay is created during data transmission.
- Problem of long chain is always high.
- Technique of selecting the cluster is not appropriate for load balancing [10].

B. PEGASIS Double Cluster Head (PDCH)

Energy efficiency is one of the important parameter of WSN. The main idea of PDCH is load balancing and to extend the network lifetime and is based on PEGASIS. In PEGASIS, there is only one cluster head in each chain but in PDCH, there are two cluster heads in each chain. It uses hierarchical structure for avoiding the long chain existing in PEGASIS [10-11].

The two cluster heads used in PDCH are referred to as bottom level cluster head and super level cluster head. The procedure of PDCH is described as follows:-

1. Hierarchy

In hierarchy structure of PDCH as shown in Fig.5, base station is at the center of circle and distance from every node to BS decides the level to which it belongs. Each node will receive the signal from base station and then according to the signal strength will detect the distance to the base station. Number of nodes, density distribution of nodes, location of base station, affects the number of levels in the hierarchy of PDCH and level has its own ID. First level ID is 0 and it belongs to BS, second level ID is 1 and it belongs to nodes closest to BS and so on.

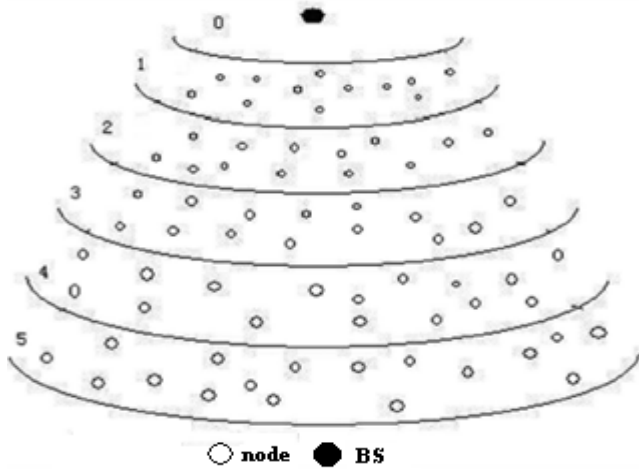


Fig. 5: Hierarchy Structure of PDCH [10]

II. Process of building chain

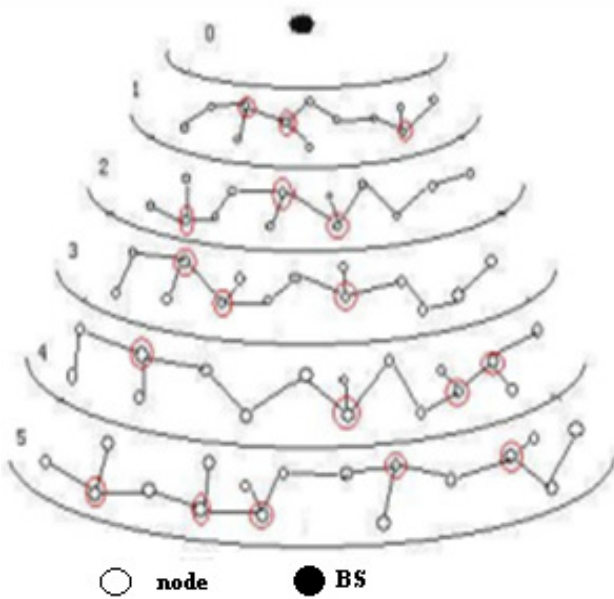


Fig. 6. Process of building chain [10]

In Fig. 6, the process of building chain is shown that if the nodes have more than one branch chain, then they have more chances to be selected as the cluster head as compared to other sensor nodes in different levels.

III. Double cluster head algorithm

There are two cluster heads in one chain as shown in Fig. 7, i.e. main cluster head and another is secondary cluster head. The function of main cluster head is data receiving and fusion and then finally transmits the data to secondary cluster head. The function of secondary cluster head is to

transmit the local level and lower level data to upper level cluster head from main cluster head.

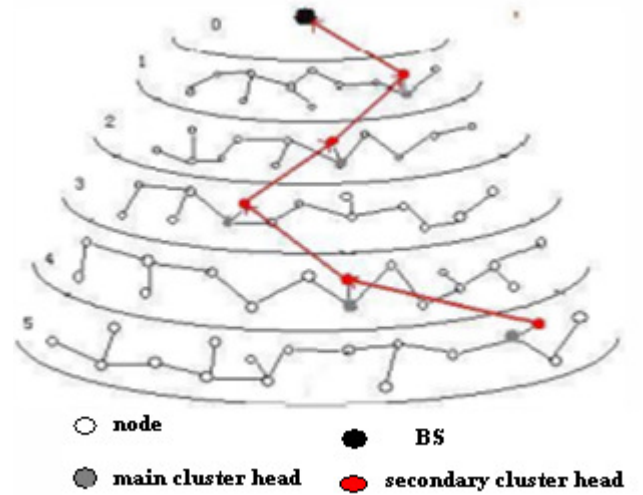


Fig. 7. Double cluster head method [10]

PDCH outperforms PEGASIS by eliminating overhead during cluster formation, minimizes the distance and limits the number of transmissions and receptions among all sensor nodes by using only one transmission to base station in each round.

C. Energy-Aware PEGASIS-Based Hierarchical Routing Protocol (EAPHRN)

EAPHRN is one of the hierarchical routing protocols for stationary WSNs. This protocol enhances the lifetime as well as the throughput of WSN. It eliminates the long chain problem of PEGASIS as shown in Fig. 8, by using new chain leader election technique. The main idea of this protocol is to make WSN optimal in terms of power consumption and to determine a low cost chain which will cover all the nodes in a network.

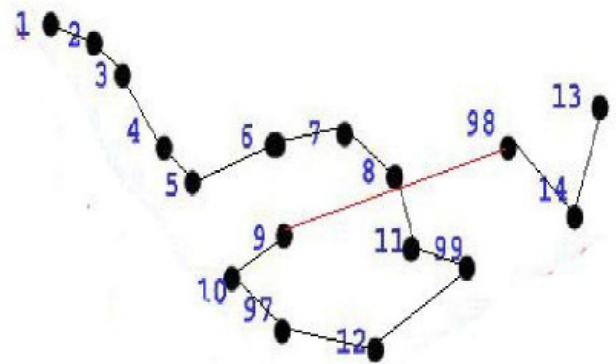


Fig. 8. Long chain problem in PEGASIS [12]

In PEGASIS greedy algorithm is used and this protocol makes an enhancement by using new algorithm that does not uses the concept of connecting to the immediate next neighbor node. EAPHRN connects to a random node which is not located far than Distance Threshold as shown in Fig.9. This protocol algorithm is categorized into two phases i.e. chain setup phase and the other is leader election method.

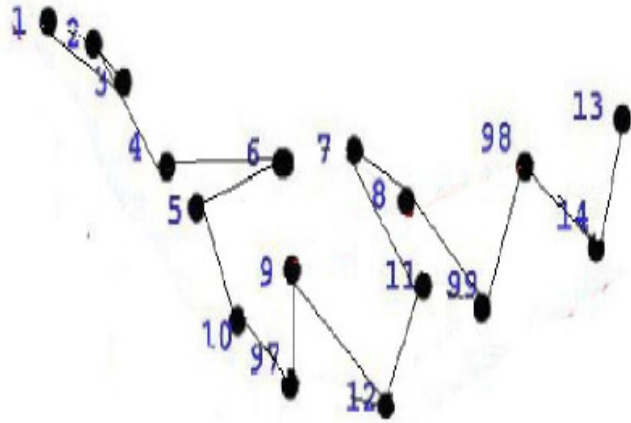


Fig. 9. EAPHRN [12]

I. Chain Setup Phase

Distance Threshold is computed before the chain is constructed. Each node computes LDT i.e. Local Distance Threshold. LDT is the average of distance between the wireless sensor node and closest n nodes.

LDT is shown in equation 1:-

$$LDT = \sum_{i=1}^n \text{dst}(i) / n \quad - \quad (1)$$

After computing LDT, DT is computed and sends it to all the sensor nodes in WSN for forming the chain.

DT is shown in equation 2:-

$$DT = \sum LDT / m \quad - \quad (2)$$

Where, m indicates the number of nodes in WSN.

II. Leader Election Method

When the chain formation is over, election of chain leader is done which is actually the closest sensor node to Base Station. Then each node senses and forwards the aggregated data to the next node in chain. In turn the node fuses its own data and then forwards to next node and so on.

Energy consumption ratio is shown in equation 3:-

$$\text{Ratio} = (\text{EnCons} / \text{EnResidual}) * 100\% \quad - \quad (3)$$

Where, EnCons refers to the amount of energy which is consumed if the sensor node is selected as a leader.

EnResidual refers to the amount if residual energy is in the sensor's node battery [12].

D. Chain Based Hierarchical Routing Protocol (CHIRON)

CHIRON is an energy efficient routing protocol which alleviates the deficiencies such as redundant transmission and data propagation delay. It is based on Beam Star concept that splits the sensing field into smaller areas which creates multiple shorter chains. The operation of CHIRON is categorized into three main phases:-

I. Group Construction Phase

This phase divides the sensing field into number of smaller areas that creates multiple shorter chains in order to reduce redundant transmission path and data propagation delay.

II. Chain Formation Phase

The nodes in each group $G_{x,y}$ are linked to form a chain $c_{x,y}$.

III. Leader Node Election Phase

A leader node is selected in each group chain for collecting and then forwarding the whole data to the base station.

IV. Data Collection and Transmission Phase

After the completion of three phases, the data is collected and transmitted. The procedure of transmission in CHIRON is same as compared to PEGASIS [4].

E. Chain Routing Protocol based on Coordinate Clustering (CRBCC)

In WSN, energy efficiency is an utmost priority to enhance the network lifetime. A two layer hierarchical routing protocol i.e. CRBCC ensures short delay and minimum energy consumption. Firstly, this protocol forms clusters along y coordinates and then by Stimulated Annealing, it selects chain leader along x coordinates. At last, it makes chain routing between the chain leaders by SA method. The procedure of CRBCC is categorized into various phases i.e. Route computation in cluster, Route computation between clusters, Route formation in cluster, Route formation between clusters and Data Gathering Phase [13].

7. CONCLUSION

PEGASIS is a routing protocol which is based on greedy algorithm but there is a problem of long chain in PEGASIS. PDCH extends the network lifetime and maintains load balancing by utilizing two cluster heads. EAPHRN is a hierarchical routing protocol that attempts to increase both the lifetime and throughput of stationary WSN. EAPHRN is based on PEGASIS and eliminates the problem of long chain

existing in PEGASIS. CHIRON is suitable for large wireless sensor networks and utilizes the Beam Star concept. CHIRON has the benefits of both chaining and clustering where the geographical area is divided into zones. CRBCC is also an hierarchical routing protocol which is not only energy efficient but also works in timely manner to meet the real time requirements.

REFERENCES

- [1] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci "A Survey on Sensor Networks" IEEE Communication Magazine, pp. 102-114, Aug 2002.
- [2] Wendi B. Heinzelman, Anantha P. Chandrakasan, Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", Proceedings of IEEE Transactions on Wireless Communications, Vol.1, No.4, pp.660-670, October 2002.
- [3] Kazem Sohrabym Daniel Minoli, Taieb Znati, "Wireless Sensor Networks, Technology, Protocols and Applications", A JohnWiley & Sons, Inc., Publication, 2007.
- [4] Kuong-Ho Chen, Taichung Jyh-Ming Huang, Chieh-Chuan Hsiao, "CHIRON: An Energy-Efficient Chain based Hierarchical Routing Protocol in Wireless Sensor Networks" Proceedings of IEEE Wireless Telecommunications Symposium, pp.1-5, 2009.
- [5] WU Cheng Dong, Cheng Fei, Ji Peng, Zhang Yunzhou, "A LEACH-Based routing protocol for wireless sensor network to optimize QoS" Proceedings of Journal of Northeastern University, Natural Sciences, Vol. 30, No.8, pp.1091-1093, 2009.
- [6] XU Dong Yi, Zhang Huazhong, "Improved Algorithm of LEACH protocol introducing residual energy", Computer Engineering and Applications, Vol.45, No. 28, pp. 115-119, 2009.
- [7] Wang Zhigang, Li Layuan, Li Chunlin, "New Uniform clustering routing protocols for wireless sensor networks", Computer Engineering and Applications, Vol. 45, No. 31, pp. 81-84, 2009,.
- [8] Wang Bo, Jiang Wei, "The Hierarchical Chain-Three Routing Protocol of improved GEGASIS" Computer Systems and Applications, pp.98-102, 2009, 2012.
- [9] Han Jingjing, Xu Zhongwei, "CCLP algorithm for improving current PEGASIS protocol", Information Technology, pp.55-58, 2008-2010.
- [10] Wang Linping, Cai Zhen, "Improved Algorithm of PEGASIS protocol introducing double cluster heads in Wireless Sensor Network", Proceedings of IEEE International Conference on Computer, Mechatronics, Control and Electronic Engineering, pp. 148-151, 2010.
- [11] Jae Duck Yu, Kyung Tae Kim, Bo Yle Jung, Hee Yong Youn, "An Energy Efficient Chain-Based Clustering Routing Protocol for Wireless Sensor Networks" Proceedings of IEEE International Conference on Advanced Information Networking and Applications Workshops, pp.383-388, 2009.
- [12] Hasan-Al-Hasan, Mohammad Qata Wneh, Azzam Sliet, Wesam Almobaideen, "EAPHRN: Energy-Aware PEGASIS-based Hierarchical Routing Protocol for Wireless Sensor Networks", Proceedings of Journal of American Science, pp. 753-758, 2011.
- [13] Liu Xiaohua, Zheng Gengsheng, "The Research of Chain Routing Protocol based on Coordinate Clustering Strategy in WSNs", Proceedings of IEEE International Conference on Information Science and Engineering, pp.1905-1908, 2009.

An Adaptive Cross Layer Routing Mechanism to Optimize Qos in MANET

V. Dhillip Kumar¹, D. Kandar², C.K. Sarkar³

¹Dept of I.T, SBCEC, Arni., ²Dept of CSE, SKPEC, T.V.Malai Dt., ³ETCE, JU, Kolkata
¹dhillipkumarit@gmail.com, ²kdebdatta@gmail.com

Abstract: This paper evaluates the cross layer routing mechanism to improve Quality of service (QoS) in MANET by combining Network layer and MAC layer protocols with Transport layer congestion control scheme to optimize the performance in Adhoc networks. MAC layer used to maintain routing table, Network layer is used in monitoring the packet data rate. These two layers helps to optimize the performance in MANET. So, combine mechanism of slow start and Arithmetic Increase mechanism of TCP helps to improve the QoS drastically. We examine the effects of different Reactive Routing protocols (AODV and DSR) with MAC Protocols used to enhance the QoS levels for MANET. MAC protocols uses distributed coordination function (DCF) and enhanced distributed coordination function (EDCF). Specifically, we access the impact of multiple wireless hops and node mobility on the throughput performance of TCP on each MAC protocol with two routing algorithms.

Similarly we examine the results in all constrained QoS parameters improvement in bandwidth-delay product, Throughput, Packet delivery ratio, and packet loss is reduced drastically to 20-25% in Enhanced DCF with AODV routing protocol in network layer and transport layer by using AIMD mechanism. But if DSR algorithm is used in the network layer instead of AODV it affects the QoS parameters. So we enhance the performance by combining cross layer architecture between proactive and reactive Protocols, hence we have taken Optimized Link State Routing (OLSR) proactive routing protocol to Optimize the QoS in MANET. The proposed OLSR routing protocol performs better than AODV and DSR during high mobility and high network load. So it can be said that, OLSR performs better than AODV and DSR at all conditions; expect very low loads when the performances are very similar.

Keywords: Mobile Adhoc networks, MAC, TCP, Slow start, AIMD, OLSR, AODV, DSR, DCF.

1. INTRODUCTION

In this paper, we evaluate the QoS parameters to optimize the MANET performance by using contention-based channel access mechanism, called EDCF to comparison with the MAC protocol IEEE 802.11 legacy. We introduced the Enhanced MAC protocol IEEE802.11e used to Enhance the QoS parameters in MANET using AODV and DSR Routing protocols. These Reactive protocols providing on-demand applications for QoS support. Based on the simulation, we compared the MAC IEEE 802.11 legacy to show that the

Enhanced Distributed Coordination Function can provide different priority traffic along with differentiated contention based mechanism. In Adhoc networks, certain QoS parameters like error rate, delay and packet loss are increased and certain parameters like throughput and delivery ratio are decreased in Transport layer is due to MAC problems and disconnection is also possible due to mobility or power failure. So, combine the mechanisms of these two layers to improve the QoS drastically so that people can design the network based on their requirements. We examine the effects of different version of MAC protocol (IEEE 802.11) with slow start and Arithmetic Increase and Multiplicative Decrease mechanism of TCP. MAC protocol uses distributed coordination function (DCF) where Enhanced version of MAC protocol uses enhanced distributed coordination function (EDCF). Specifically, the interaction between transport layer and the MAC protocol has a significant impact on the achievable throughput, Packet Delivery Ratio, Bandwidth Delay Product and packet loss in ad hoc networks.

ARCHITECTURAL DESIGN

Application Layer (This layer generate multimedia packets and assign priority)
Transport layer (Implement Slow start & AIMD mechanism)
Network layer (Implement OLSR Protocol)
MAC layer (Implement EDCF&DCF algorithm)

Fig. 1. Layered structure

2. PROBLEM DESCRIPTION

The introduction of Multimedia services like real-time audio, video and data applications to overcome a drawback of on-demand technical services in wireless networks. Traditional QoS Routing protocols like Resource reservation protocol (RSVP) cannot be easily changed to the mobile environment due to the error-rate and mobility nature of wireless links for data stream applications. This is especially

true for Mobile Ad Hoc Networks (MANETs) where every node moves arbitrarily causing the multi-hop network topology to change randomly and at unpredictable times. In order to prove its correctness and efficiency the system is implemented and simulated using the ns-2 network simulator.

3. EXISTING SYSTEM

OSI architecture allows each layer to be extracted independently, which simplifies the implementation of the architecture. However, the strictly layered architecture might not be the best model, because often times it is not possible to optimize the network performance according to different situations without interaction among the different layers. Thus a cross-layer design is the solution for enhancing QoS in various wireless networks such as sensor networks, cellular networks and ad hoc networks. A cross-layer design is used to adjust the transmission rate in transport layer and provide channel information to the MAC layer from the physical layer; resource allocation is determined in the MAC layer according to multi-path routing information from the network layer. Our previous work was the interaction between transport layer mechanisms and MAC protocols and discussed the performance improvement between AIMD, Slow start with IEEE 802.11e and AIMD, Slow start with IEEE 802.11. Disadvantages of existing system, It is not possible to optimize the network performance according to different situations without interaction among the different layers. Network layer not able to detect the path to deliver the packets, so there is no guarantee in the QoS In MANET,

4. PROPOSED SYSTEM

The proposed works evaluates the performance evaluation by combining Network layer and MAC layer protocols with Transport layer congestion control mechanisms operating in a mobile adhoc network. In Adhoc networks, certain QoS parameters like error rate, delay and packet loss are increased and certain parameters like throughput and delivery ratio are decreased in Transport layer is due to MAC problems and disconnection is also possible due to mobility because the network layer is not able to detect the path to deliver the packets. So, combine the mechanisms of these three layers to improve the QoS drastically. Advantages of proposed system, Optimized State of Routing, Enhancement in the quality of service, Decreased error rate, delay, packet loss, increased throughput and delivery ratio.

OLSR (Optimized Link State Routing)

The OLSR is a link state proactive protocol used to exchange topology information with neighbor nodes of the mobile ad-hoc network. OLSR based on multipoint relays to minimize the packet flooding. During the flooding process

packets will be forwarded to selected nodes. The control messages Hello and Topology Control used to sense the neighbor nodes to broadcast the forwarded messages and then collecting information from each protocol evaluating the link state information to improve the performance parameters in MANET. Sending and receiving packets of other nodes running a different protocol used to compute the hop count information from source and destination nodes to improve the network reliability using shortest path mechanism to forwarding packets. Advantages: OLSR is also an IP based routing protocol, it does not need centralised server to handle its routing process, being a proactive protocol, and routes to all destinations within the network are known and maintained before use.

Slow-Start

Slow-start is the process of congestion avoidance strategy used by TCP for data transmission control. Slow-start mainly depends on average throughput of a TCP connection in terms of the packet loss probability, the packet size, and the round-trip time. The response function of TCP is used to avoid sending more data packets to prevent from network congestion.

Fast recovery

Fast retransmit and fast recovery are usually implemented together to use of congestion avoidance, fast recovery is similar to slow-start algorithm. When duplicate acknowledgement received it will resend the lost segment immediately which uses fast retransmit followed by congestion avoidance. the congestion window size is reduced due to fast recovery mechanism.

Additive Increase and Multiplicative Decrease (AIMD)

The additive increase/multiplicative-decrease (AIMD) algorithm with implicit loss feedback to TCP Congestion Avoidance. In wireless networks failure to distinguish congestion loss from corruption loss. AIMD congestion window used to reduce the exponential time when traffic takes place. This mechanism used to increase the window size for improve the transmission rate, available bandwidth, until packet loss occurs. If no congestion detected, periodically increase the data rate, if congestion detected, immediately decrease the data until a loss is detected, Due to this reason Congestion window is increased by 1 maximum segment size in every round trip time (RTT).

5. IMPLEMENTATION

The OLSR protocol proposed in this project is compared with the AODV and DSR protocol. NS2 is an open source method and program to implemented (e.g. Linux,). It is a discrete event time simulator. With, a single entity can simulate several network nodes in the system. Network

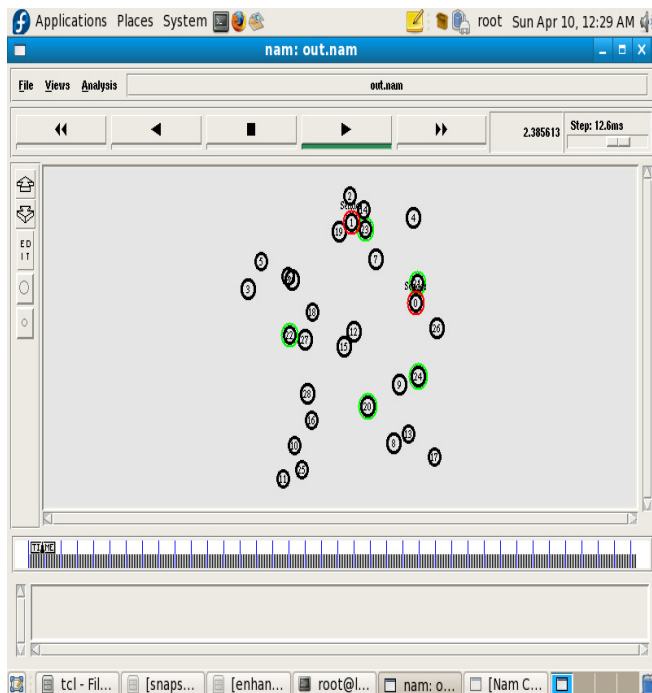
Simulator (NS2) is used to simulate these protocols. Second and the more important reason are related to our proposed protocol, working on the top of well known protocol such as OLSR protocol. NS2 protocol has many routing Protocols. (e.g.: DSR, AODV). We implement a proactive routing protocol called OLSR (OPTIMIZED LINK STATE ROUTING PROTOCOL) by inserting the network layer and then recompiled the Network Simulator (NS2).

Simulation Parameters

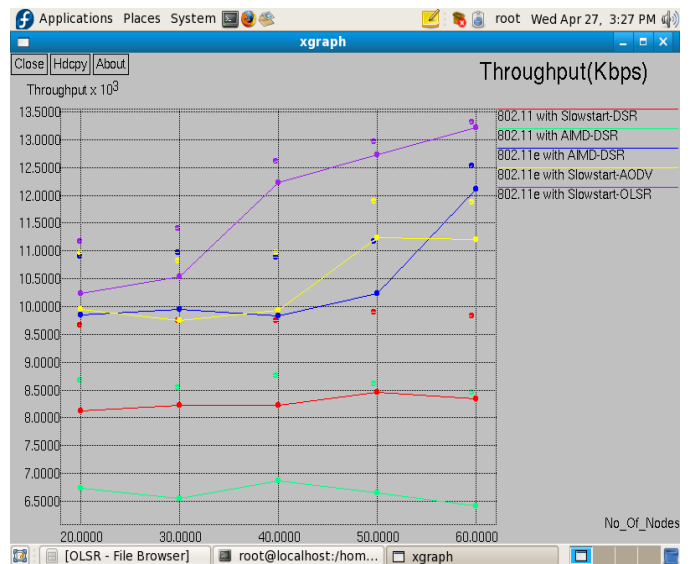
Parameters	Values
Routing Protocol	OLSR
MAC	802.11
Transmission Rate	1 Mbps
Simulation Time	100 seconds
Mobility	RWP
No. of nodes	20
Node speed	2 m/s
Pause time	0, 30s
Traffic	CBR
Network Area	1500 m X 1500m

6. SIMULATION RESULTS

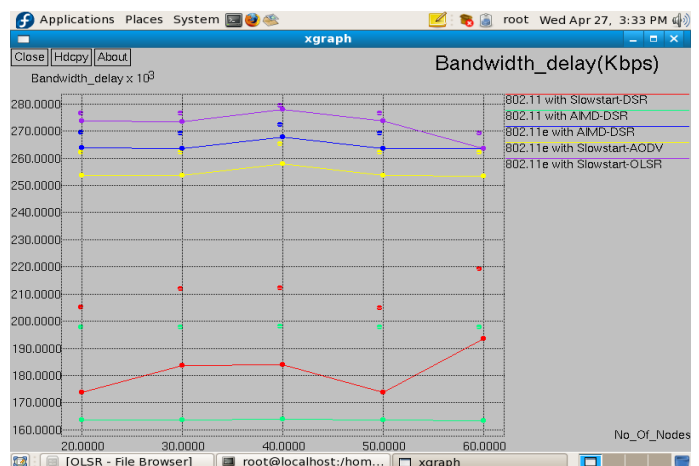
Route Discovery



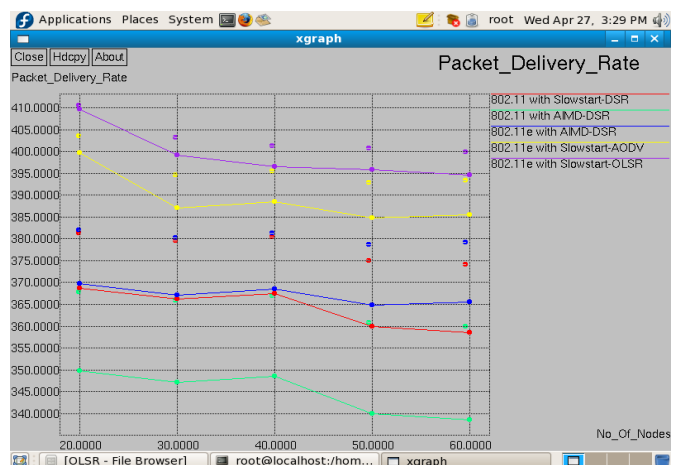
Throughput



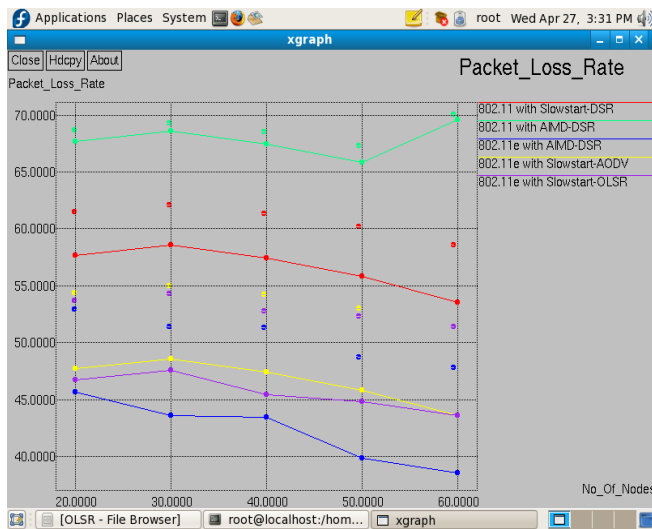
Bandwidth_Delay



Packet_Delivery_Rate



Packet Loss Rate



7. CONCLUSION

The proposed OLSR routing protocol performs better than AODV and DSR during high mobility and high network load. OLSR always maintains a routing table; most often it can provide some routes quickly. So the average delay is reduced significantly. Packet delivery ratio is improved as it maintains the QoS information and looks for a path satisfying the QoS requirements of the applicants. Moreover it sends smaller number of control packet to handle route discovery and route failure. As a result, the control overhead is reduced. The trade-off is that each node requires more memory to store the neighbor information and comparatively larger routing table. The nodes also need more processing power to manipulate the neighborhood information and calculate the routes based on QoS information. So it can be said that, OLSR performs better than AODV and DSR at all conditions; expect very low loads when the performances are very similar. So, MAC layer information can be used to construct the neighbor table which will allow enhancing the performance in terms of delay and packet delivery as well as minimizing the control overhead. The results show that the interaction between transport layer with the Network and MAC protocols has a significant impact on the achievable throughput, Packet Delivery Ratio, Bandwidth Delay

Product and packet loss in Adhoc networks. These three layers know the status of other layers and collectively improve the QoS performance in Adhoc networks. The OLSR protocol enhances the quality of service.

REFERENCES

- [1] J.Premalatha, "Enhancing QoS in MANETS by Effective Routing", KEC, Erode, India IEEE (2010).
- [2] Amina Akhter, "Modified AODV for Multi-constrained QoS Routing and Performance Optimization in MANET" at 2009.
- [3] Nur Idawati Md Enzai, Farhat Anwar, Omer Mahmoud, "Evaluation Study of QoS-Enabled AODV," Proceedings of the ICCCE, Malaysia, 2008.
- [4] Hongxia Sun, Herman D. Hughes, "Adaptive QoS Routing by Cross-Layer Cooperation in Ad Hoc Networks", EURASIP Journal on WSN 2005, 661–671.
- [5] Nityananda sarma, sukumar Nandi, Rakesh Tripathi, "Enhancing Route Recovery for QAODV Routing in mobile Adhoc Network", The international Symposium on parallel Architecture, Algorithm and Network. DOI 10.11.09/ I-SPAN 2008.
- [6] Choi S (2003), 'IEEE 802.11e contention based channel access with Enhanced DCF performance evaluation' Proc. IEEE ICC.
- [7] Maarten Hoebe and Menzo Wentink, "Enhanced QoS through Virtual DCF," IEEE 802.11-00/3 51, October 2000.
- [8] Paolo Giacomazzi (2006), 'Quality of service for packet telephony over Mobile Ad Hoc Networks'
- [9] Qixiang Pang, Soung C.Liew (2005), 'Design of an effective loss distinguishable MAC protocol for 802.11 WLAN'.
- [10] Lei Chen "Protocols for Supporting Quality of Service in Mobile AdHoc Networks," dissertation, University of Rochester Rochester, New York, 2006.
- [11] Schiller J (2003), 'Mobile communication', Pearson education private limited, Singapore.
- [12] Menzo Wentink, saishankar Maarten Hoebe and Stefan Mangold, "Multiple Frame Exchanges during IEEE802.11e EDCF, TXOP," IEEE, Jan 2002.
- [13] Xiao Y (2005), "Performance analysis of priority schemes for MAC IEEE 802.11 and IEEE 802.11e Wireless LANs', IEEE Transaction on wireless communication Vol 4.
- [14] Yang Xiao, (2004), 'IEEE 802.11e.QoS Provisioning at the MAC layer' IEEE wireless communications, 1536- 1284, pp. 72-79.
- [15] C. Zhu and M. Corson, "QoS Routing for Mobile Ad Hoc Networks," in *IEEE Informcom*, 2002.

A Study on Future Advancements in Security for MANET

Samta Suman Lodhi¹, Radhey Shyam Lodhi²

^{1,2}2709 West Royal Lane, Irving, Texas(U.S.A)
¹samtalodhi@gmail.com, ²rs_lodhi@rediffmail.com

Abstract: In this paper, we've dealt with the primary challenge of building security in MANETs and also maintain the information required to properly route traffic. Along with this, we have also discussed about the potentials of 4G technology.

Keywords: MANET, Security, WIMAX, LTE, 4G, Internet Protocol.

1. INTRODUCTION

Traditionally, the service provision in 2G networks, e.g. GSM, has been mainly based on voice services, closed business model support and limited operator differentiation due to a narrow set of offered services. Actually, mobile service provision is facing important advancements towards more flexible business models, with the introduction of new 2.5G/3G generations of mobile communication systems, like GPRS, UMTS and CDMA2000. Unfortunately, these 2.5/3G networks entail limitations to fulfill requirements imposed by current mobile users specially with the "anytime, anywhere with anybody" type of communication. Since 1970s, the research of ad hoc networking was mainly large scale networks for emergency/rescue and military purposes respectively for disaster and battlefield communication applications. *Large scale isolated ad hoc networks* are not suited to transport a large amount of data due to their very low traffic performance, slow topology convergence and security problems. However, these could be used to transport very urgent short messages (e.g. to inform about the location of an accident or to transmit tactical commands). Since 1990s, *small isolated ad hoc networking* has been experiencing a growing interest in the commercial and residential areas due the proliferation of small information computational devices and the emerging wireless technologies (IEEE 802.11, Bluetooth). This development is driven by the need to exchange digital information among people in direct contact enabled by ad hoc networking among a number of wireless nodes. Small In the context of the heterogeneous and integrated 4G environment, ad hoc networking is considered an important solution to extend the radio coverage of wireless systems and multimedia Internet services to wireless environments [1]-[2]. In these *integrated ad hoc networks* mobile ad hoc hosts and routers can gain Third/fourth generation cellular networks (3G/4G) are broadband wireless mobile networks that has evolved from

the 1st to the 2nd and 3rd generation networks. The still evolving 4th generation network is expected to be deployed in later 2011.

2. THE NEXT GENERATION TECHNOLOGY

4G is short for Fourth (4th) Generation Technology. 4G Technology is basically the extension in the 3G technology with more bandwidth and service offers in the 3G. The expectation for the 4G technology is basically the high quality audio/video streaming over end to end Internet Protocol. If the Internet Protocol (IP) multimedia sub-system movement achieves what it going to do, nothing of this possibly will matter. 4G is intended to provide:

- High speed
- High capacity
- Low cost per bit
- IP based services for video, data and voice (VoIP).

4G is all about integrated, global network that is based on an open system approach. At the moment we have several technologies each capable of performing some of functions like broadband data access in mobile or nomadic environment, supporting voice traffic using voice over IP etc[3]. But what we really need is a deployment of new technologies that allow merging, bridging and integrating all these repeated system into an information delivery system of the twenty first century. 4G stands for Fourth Generation and is the latest technology with high speed transferability of data with security measurements. It is coming with wireless broadband for the instant download. Talking about the standard of 4G technology, so far, two technologies are supposed to be the features of 4G.

A. WiMAX

WiMAX stands for Worldwide Interoperability of Microwave Access previously worked as fixed wireless facility under the 802.16e band. Now the modified standard 802.16m has been developed with the properties of speed, wide spectrum, and increase bandwidth. 4G has an

advantage of having the WiMAX as a product because IEEE has introduced and released it already, therefore its economic as there is no need to pay for its manufacturing price. 4G supports two basic equipments; WiMAX Network system (network infrastructure) and mobile phone set. Smartphones with Wireless Access introduced in the market are the model 4G mobiles. These smartphones are equipped with the wireless internet accessibility and there is no fear of losing connection while travel from one tower to another tower range. WiMAX or mobile structural design will become progressively more translucent, and therefore the acceptance of several architectures by a particular network operator ever more common.

B. LTE

Parallel to WiMAX, LTE (Long Term Evolution) is introduced by Verizon. LTE is considered to be promising high data transfer speed. LTE is supposed to provide internet facility using both systems. It has the ability of transition from one mode to another. LTE is developed on radio waves technology. This not only increases the speed but also the amount of data allowed through the same bandwidth and results into lower cost. As LTE is compatible with 3G technology so, it not only increases the speed but also prevents the need of new network and can work through the same infrastructure. LTE will not only support the functions of 3G but also incorporate some newer ones. LTE is using MIMO (Multiple input multiple output) and is able to send and receive huge data. It is negative in the sense that it will overload the base stations networks. 4G Technology offers high data rates that will generate new trends for the market and prospects for established as well as for new telecommunication businesses. 4G networks, when tied together with mobile phones with in-built higher resolution digital cameras and also High Definition capabilities will facilitate video blogs. After successful implementation, 4G technology is likely to enable ubiquitous computing, that will simultaneously connect to numerous high data speed networks offering faultless handoffs all over the geographical regions. Many network operators possibly utilize technologies for example; wireless mesh networks and cognitive radio network to guarantee secure connection & competently allocates equally network traffic and bandwidth.

3. SECURITY ADVANCEMENTS IN MANET WITH 4G

A mobile ad hoc network (MANET), sometimes called a mobile mesh network, is a self-configuring network of mobile devices connected by wireless links. Each device in a MANET is free to move independently in any direction, and will therefore change its links to other devices frequently. Each must forward traffic unrelated to its own use, and therefore be a router. The primary challenge in building a MANET is equipping each device to continuously maintain

the information required to properly route traffic. Authentication research has determined that for a positive identification, elements from at least two, and preferably all three, factors be verified. The three factors (classes) and some of elements of each factor are:

A. The Ownership Factors:

Something the user has (e.g., wrist band, ID card, security token, software token, phone, or cell phone).

B. The Knowledge Factors:

Something the user knows (e.g., a password, pass phrase, or personal identification number (PIN), challenge response (the user must answer a question)).

C. The Inherence Factors:

Something the user is or does (e.g., fingerprint, retinal pattern, DNA sequence (there are assorted definitions of what is sufficient), signature, face, voice, unique bio-electric signals, or other biometric identifier).

At small scale, the identity verification can be managed by the nodes themselves, as handshaking by virtue of their proximity [4], but at relatively larger scale it becomes complex and the nodes identity verification demands the authentication involvement of TTP [5]. There are schemes that are based on the concept of self-organization in MANETS [6] thoroughly without TTP connection where the identity is resolved by the nodes themselves or some hybrid form of above two schemes might be used [4]. The 4G system was originally envisioned by the Defense Advanced Research Projects Agency. The DARPA selected the distributed architecture, end-to-end Internet protocol (IP), and believed at an early stage in peer-to-peer networking in which every mobile device would be both a transceiver and a router for other devices in the network eliminating the spoke-and-hub weakness of 2G and 3G cellular systems. Since the 2.5G GPRS system, cellular systems have provided dual infrastructures: packet switched nodes for data services, and circuit switched nodes for voice calls. In 3g and 4G systems, the circuits switched infrastructure is abandoned, and only a packets witched network is provided. This means that traditional voice calls are replaced by IP telephony. Cellular systems such as 4G allow seamless mobility; thus a file transfer is not interrupted in case a terminal moves from one cell (one base station coverage area to another, but handover is carried out. The terminal also keeps the same IP address while moving, meaning that a mobile server is reachable as long as it is within the coverage area of any server. In 4G systems this mobility is provided by the mobile IP protocol, part of IP version 6, and while in earlier cellular generations it was only provided by physical layer and data link layer protocols. In addition to seamless mobility, 4G provides flexible interoperability of

the various kinds of existing wireless networks, such as satellite, cellular wireless, WLAN, PAN and systems for accessing fixed wireless networks. 4G stands for the fourth-generation cellular network. Although it is generally agreed that 4G is going to offer better communication technology than 3G, it is still undefined as to which areas should be really improved upon, and in which ways, from 3G. Researchers are often pointing towards integration whereas business institutions are working on upcoming technologies that will make 4G more attractive to the business community by implementing it more customer-friendly. New support for mobility is the primary concern of Hussian *et. al.* [7] and they pointed out insufficient 3G mobility constrained by bandwidth that should be significantly increased. According to them, the significant progress that 4G can achieve in the area of mobility is unifying different and currently separated environments into a single fixed OWA (Open Wireless Architecture) that will achieve high connectivity by accessing all kinds of networks. Providing single terminal that will effectively access the best available internet connection will increase and speed up device usability under 4G. Integration is the key concept in defining 4G capabilities since we should support all kinds of multimedia by offering single access to all wireless networks. Understanding the significance of unifying Wi-Fi, WiMax and Cellular networks into one product, Woerner and Howlader proposed that the most important factor of 4G will be “seamless integration of wireless networks” based on flexibility of the software radio technology, with improved bandwidth capacity, and improved routing techniques allowing multi-hop peer-to-peer networks. Due to the lack of single military scenario where and how 4G will be used, it is critical that future wireless technology will be capable of effortlessly accessing all kind of radio communications. Bauer *et. al.* addressed that enhanced cellular range and capacity, supported by Wi-Fi and WiMAX networks is the vision of 4G. However, considering the fast development of WiMAX networks, and the increasing range of Wi-Fi standards, they argue that these new wireless networks can in the future substitute cellular networks such as the current 3G. They also addressed that it is “misleading” calling the evolution of cellular technology in terms of generations because this would “suggest a linear progression” which is not the case. Finally, they also evaluated business opportunities of 4G pointing out on establishing a global standard, along with open architecture, and supporting multiple interfaces all over the world, as the keys to economic success. Steer [7] addressed 4G is officially designated by IEEE as “Beyond 3G.” Characterized by wireless broadband with over 100Mbps mobile capacity and 1Gbps stationary bandwidth supported by OFDM, MIMO, and software defined radio, Steer presents new 3G’s components that will upgrade it up to 4G. The idea of upgrading 4G is shared by two other groups working on the next generation technology 3GPP and 3GPP2 developing new versions of UMTS and CDMA2000 cellular systems respectively. After introducing HSDPA

(High-Speed Downlink Packet Access) in release 5, HSUPA (High-Speed Uplink Packet Access) in release 6, and HSOPA (High Speed OFDM Packet Access) in release 7, the 3GPP group project is working on release 8 – the UMTS (Universal Mobile Telecommunications System) Revision 8 LTE (Long Term Evolution) that will introduce 4G on UMTS foundations. The 3GPP plans presented in Technical Report (TR) 25.913 that are going to be concluded in September 2007 [6] expects cell coverage between 5 to 30 km, latency below 100ms, 100 Mbps/50Mbps downlink/uplink data rate within 20MHz spectrum allocation, high performance mobility up to 120km/h that between networks can be increased as much as up to 500km/h. The same report signifies the importance of IPbased networks with support of MIMO and OFDMA. The Pioneer and Inventor of 3G/WiFi Convergence Systems and Technologies, Top Global USA, Inc. created the first such 4G picture, the first mobile router that links 3G/4G Cellular and Wi-Fi networks. Providing seamless routing and secure connectivity, Top Global’s router maintains connection in moving vehicles with 802.11n, HSDPA, and WiMAX wireless access points simultaneously.

4. CONCLUSION

A mobile ad hoc network (MANET) is a temporary, self-organizing network of wireless mobile nodes without the support of any existing infrastructure that may be readily available on the conventional networks. Since there is no fixed infrastructure available for MANET with nodes being mobile, routing becomes a very important issue. In addition, we also explained the various emerging applications and future trends of MANET.

At the moment we have several technologies like 2G, 3G etc. each capable of performing some of functions like broadband data access in mobile or nomadic environment, supporting voice traffic using voice over IP etc. But what we really need is a deployment of new technologies that allow merging, bridging and integrating all these repeated system into an information delivery system of the twenty first century. 4G stands for fourth Generation cellular network. Nowadays, network technology plays a significant role on science and business area. Everyday new technologies are emerging. Fourth Generation (4G) is the next generation of wireless networks that will replace third Generation (3G) networks sometimes in future.

REFERENCES

- [1] B.Xu, S. Hischke and B. Walke. "The Role of Ad hoc Networking in Future Wireless Communications". In Proc. ICCT. Beijing, 2003.
- [2] Mobile IP-based Network Developments (IST-2000-28584 MIND). Project homepage: <http://www.ist-mind.org>

-
- [3] F. Bader, C. Pinart, C. Christophi, E. Tsiakkouri, I. Ganchev, V. Friderikos, C. Bohoris, L. Correia, L. Ferreira. "User-Centric Analysis of Perceived QoS in 4G IP Mobile/Wireless Networks". PIMRC'2003, Pp. x.1-x.7, 7-10 September 2003. Beijing, China. ISBN 0-7803-7823-7.
 - [4] F. Stajano and R. J. Anderson. —The resurrecting duckling: Security issues for ad-hoc wireless networks. In 7th Security Protocols Workshop, vol. 1796 of LNCS, UK, 1999. Springer-Verlag, Germany
 - [5] Yuh-Min Tseng. —A heterogeneous-network aided public-key management scheme for mobile ad hoc networks, Published on 10 February 2006 in Wiley InterScience, Int. J. Network Mgmt; 17: 3–15
 - [6] S. Capkun, L. Buttyan and J-P Hubaux. "Self- Organized Public-Key Management for Mobile Ad Hoc Networks ", IEEE Transactions on Mobile Computing, Vol. 2, No. 1, Jan-Mar 2003, pp. 52-64
 - [7] Hussian S., Hamid Z., and Khattak N., "Mobility Management Challenges and Issues in 4G Heterogeneous Networks."

A Comparative Study of Security Attacks in Bluetooth, Wi-Fi and Wimax

Nandini Deb¹, Tushar Saxena², Himanshu Goyal³

^{1,2,3}M.Tech 1st Year, Dept. of TSE, AITTM, Amity University, Noida, India

¹nandini.deb2011@gmail.com, ²tusharsaxena10@gmail.com, ³goyalhimanshu19@gmail.com

Abstract: The main aim of this paper is to discuss the security features and weakness of IEEE 802.15, IEEE802.11 and IEEE802.16 i.e. Wireless PAN, Wireless LAN and Wireless WAN networks. Bluetooth, Wi-Fi and Wi-Max are the three most widely used wireless networking technologies. These networks are exposed to many types of risks and have various flaws in their respective protocol structure. Bluetooth belongs to a category of Short-range Wireless technologies, Wireless LAN or Wi-Fi is the LAN network we use in our home, offices or buildings etc to provide user the network availability whereas WI-MAX is used on a bigger scale MAN i.e. providing the network coverage on a metro scale through cellular networks to compensate the wired broadband services. In this paper, we provide a study of these popular wireless communication standards in current scenario, evaluating their security features in terms of various metrics. This paper focuses on the various types of attacks on these networks and the countermeasures to overcome them.

IndexTerms: Wi-Fi, WiMax, Bluetooth, Security, Countermeasures.

1. INTRODUCTION

The wireless local area network (WLAN) is today often taken as a default interface for networked devices by users and manufacturers alike. Wireless networking is an essential productivity tool for today's mobile workforce. But this was not the picture even 15 years back. The WLAN of that day appeared to lack both the throughput of the wired local area network (such as 10/100 Ethernet LAN) and the coverage of the cellular network[1]. The WLAN to that point had largely evolved as a slow and unreliable emulation of the wired LAN, only without the wire.

In just 25 years wireless technologies have changed the face of networking and have also changed the concept of network topology which has been possible due to the relentless progress in silicon technology with higher integration, lower costs, more capabilities and technical advances in air interfaces i.e. higher efficiency for voice and data services, lower infrastructure capital costs. Wireless networks can be classified broadly as *Wireless Personal-Area Networks* (WPAN), *Wireless LANs* (WLANs), and *Wireless Wide-Area Networks* (WWANs). WPANs operate in the range of a few feet, whereas WLANs operate in the range of a few

hundred feet and WWANs beyond that. In fact, wireless WANs can operate in a wide range—a metropolitan area, cellular hierarchy, or even on intercity links through microwave relays. Wireless topologies include: point to point, line of sight, scatter and reflective, cellular, radio broadcast.

Table I. Wireless technology comparison chart

	Bluetooth	Wifi(a)	WiFi(b)	WiFi(g)	WiMAX
STANDARD	802.15	802.11a	802.11b	802.11g	802.16
FREQUENCY	2.45 GHz	5 GHz	2.4 GHz	2.4 GHz	2-66 GHz
SPEED(Mbps)	0.72	54	11	54	80
RANGE	10 m	50 m	100 m	100 m	50 km
PROS	Low Cost	Speed	Low Cost	Speed	Speed, Range
CONS	Range	Cost	Speed	Cost, Range	Cost

2. GENERAL VULNERABILITIES AND THREATS IN WIRELESS NETWORKS

There are various attacks that can cause compromise to one or more of the three fundamental security objectives of integrity, confidentiality and availability. The genera threats to the wireless networks are as follows:

A. Malicious Association

“Malicious associations” are when wireless devices can be actively made by attackers to connect to a company network through their cracking laptop instead of a company access point (AP). These types of laptops are known as “soft APs” and are created when a cyber criminal runs some software that makes his/her wireless network card look like a legitimate access point[2]. Once the thief has gained access, he can steal passwords, launch attacks on the wired network, or plant trojans. Since wireless networks operate at the Layer 2 level, Layer 3 protections such as network authentication and virtual private networks (VPNs) offer no barrier. Wireless 802.1x authentications do help with some protection but are still vulnerable to cracking.

B. Accidental association

When a user turns on a computer and it connects to a wireless access point from a neighboring company's overlapped network, the user may not even know that this association has occurred, called as accidental association.

C. Ad-Hoc Networks & Pan Networks

Ad-hoc networks are peer-to-peer networks between wireless computers that do not have an access point in between them. These networks can pose a security threat unless sufficient encryption methods are used to provide security.

Bluetooth devices like barcode readers, PDAs, wireless keyboards and printers are not safe from cracking and should be secured.

D. Mac Spoofing

MAC spoofing is a technique for changing a factory-assigned Media Access Control (MAC) address of a network interface on a networked device. The changing of the assigned MAC address may allow the bypassing of access control lists on servers or routers, either hiding a computer on a network or allowing it to impersonate another network device. A user may wish to legitimately spoof the MAC address of a previous hardware device in order to reacquire connectivity after hardware failure.

E. Man-in-The-Middle Attacks

The man-in-the-middle attack in cryptography an computer security is a form of active eavesdropping in which the attacker makes independent connections with the victims and relays messages between them, making them believe that they are talking directly to each other over a private connection, when in fact the entire conversation is controlled by the attacker. This attack can succeed only when the attacker can impersonate each endpoint to the satisfaction of the other-it is an attack on mutual authentication.

F. Caffè Latte Attack

The Café Latte attack allows you to obtain a WEP key from a client system. Briefly, this is done by capturing an ARP packet from the client, manipulating it and then sends it back to the client. The client in turn generates packets which can be captured by airodumping. The attacker uses the ARP responses to obtain the WEP key in less than 6 minutes.

G. Network Injection

In a network injection attack, a cracker can make use of access points that are exposed to non-filtered network traffic, specifically broadcasting network traffic such as "Spanning

Tree" (802.1D), OSPF, RIP, and HSRP. The cracker injects bogus networking re-configuration commands that affect routers, switches, and intelligent hubs[3]. A whole network can be brought down in this manner and require rebooting or even reprogramming of all intelligent networking devices.

H. Denial of Service

DOS is a type of attack on a network that is designed to bring the network to its knees by flooding it with useless traffic. Many DoS attacks, such as the *Ping of Death* and *Teardrop* attacks, exploit limitations in the TCP/IP protocols. For all known DoS attacks, there are software fixes that system administrators can install to limit the damage caused by the attacks. But, like viruses, new DoS attacks are constantly being dreamed up by hackers.

3. ATTACKS ON BLUETOOTH AND COUNTERMEASURES

A. Bluesnarfing

Bluesnarfing attacks involve a hacker covertly gaining access to your Bluetooth-enabled device for the purpose of retrieving information, including addresses, calendar information or even the device's International Mobile Equipment Identity[4]. With the IMEI a hacker could route one's incoming calls to his cell phone. Bluesnarfing was a bigger problem on cell phones between 2003 and 2004. It is hard to do, and the necessary software can be tough to obtain.

SOLUTIONS: Firmware updates have reduced the threat considerably. In addition, placing your phone in a nondiscoverable mode makes it harder on the attacker, because he then needs additional software to locate your Bluetooth-signal.

B. Bluebugging

Bluebugging means hacking into a Bluetooth device and using the commands of that device without notifying or alerting the user. By bluebugging, a hacker could eavesdrop on phone conversations, place phone calls, send and receive text messages, and even connect to the Internet. Bluebugging exploits a different vulnerability than bluesnarfing. It's a firmware issue commonly associated with older cell phones.

C. Bluejacking

Bluetooth devices have the ability to send so-called wireless business cards. A recent trend has been to send anonymous business cards with offensive messages, and frankly, it's easy to do. But it doesn't put data in jeopardy. Bluejacking requires an attacker to be within 10 meters of a device.

SOLUTIONS: If someone bluejacks a person, he/she could probably see his face. Bluejack messages should never be added to the contacts list. And to avoid the nuisance altogether, the phone should be kept on non discoverable mode.

D) Denial of service

DOS attacks occur when an attacker uses his Bluetooth device to repeatedly request pairing with the victim's device[5]. Unlike on the Internet, where this type of constant request can bring down services, a Bluetooth DOS attack is mostly just a nuisance, since no information can be transferred, copied or attained by the attacker.

SOLUTIONS: DOS attacks are the easiest to perform and can drain a device's battery or temporarily paralyze the phone or PDA. However, since this attack relies on the proximity of the attacker to the victim, it's easy to stop. Just walk away. Currently, there are few software defenses against this type of assault.

E. Car Whispering

This type of attack uses a software and it results in transmission and reception of audio to and from Bluetooth enabled car audio system. Attacker would be able to add or announce something he wants to and would also be able to listen to the conversation going inside.

F. Fuzzing attacks

It includes sending and transmitting malformed data to a device's Bluetooth radio and monitors the functionality of the device. If the device functionality is slowed down after this attack then there is a serious vulnerability that has occurred in the security protocol stack.

Solutions

In order to stand against the vulnerabilities and attacks, the Bluetooth device manufacturers are upgrading the security standards. Manufacturers should not use the standard passkeys in their products and there should be some direct interaction with the device that allow device to connect. Also the handsfree unit should be switched to invisible mode when there is no authorized connection is running for the time.

4. ATTACKS ON WI-FI AND COUNTERMEASURES

A. Rogue access points

Unsanctioned, unknown and unmanaged devices inside the network become wide-open back doors, providing easy routes for malware to come in and information to leave the network. The first step in countering this problem is to

enforce no-wireless zones, ensuring that access points do not appear where they are not allowed. Banning wireless access completely has been the typical first reaction to this problem in most agencies. In some sensitive areas, such as military networks and the Federal Aviation Administration, they still tend to have no-wireless zones. However, over the last six or seven years, the trend is toward wireless as almost everyone demands it. The problems with rogue access do not stop there. After administrators have the policies and tools in place to manage approved access points, rigorous monitoring is necessary.

B. Misconfiguration

Misconfiguration of switches and access points still represents a major problem because wireless is a new technology, and administrators have less experience with it than with wired networks. As with most other equipment, default settings often are a no-no, and devices need to be tuned to conform to policies and best practices.

C. Unmanaged use of wireless technology outside the enterprise

More and more employees are becoming mobile, using devices on outside, open networks. That can leave them vulnerable to malicious traffic. That is especially true with Windows 7 support for Virtual Wi-Fi, which allows neighbours to share access to a laptop. Without Windows 7, laptops typically act only as a client on wireless networks. But Virtual Wi-Fi allows the client to also act as an access point and provide services for other clients, creating ad hoc, peer-to-peer networks that can put users at risk.

D. Hackers

Active attacks on wireless links are a growing problem as mobile and wireless computing offers increasingly attractive targets to hackers. After a device becomes powerful enough and the information they contain becomes valuable enough, they attract the attention of bad guys and are likely to fall victim to exploits. A good defence against hackers is educational and technical knowledge evolution. More enterprises are realizing they need to have a 24/7 monitoring system for wireless. As adoption increases, various sensitive markets, such as the Defence Department and payment card industry, are becoming more prescriptive in their security, with requirements for best practices in procuring and managing the technology.

E. Denial of Service

In this kind of attack, the intruder floods the network with either valid or invalid messages affecting the availability of the network resources. Due to the nature of the radio transmission, the WLAN are very vulnerable against denial of service attacks. The relatively low bit rates of WLAN can

easily be overwhelmed and leave them open to denial of service attacks. By using a powerful enough transceiver, radio interference can easily be generated that would enable WLAN to communicate using radio path.

F. Spoofing and Session Hijacking

This is where the attacker could gain access to privileged data and resources in the network by assuming the identity of a valid user. This happens because 802.11 networks do not authenticate the source address, which is Medium Access Control (MAC) address of the frames. Attackers may therefore spoof MAC addresses and hijack sessions. Moreover, 802.11 do not require an Access Point to prove it is actually an AP. This facilitates attackers who may masquerade as AP's. In eliminating spoofing, proper authentication and access control mechanisms need to be placed in the WLAN.

G. Eavesdropping

This involves attack against the confidentiality of the data that is being transmitted across the network. By their nature, wireless LANs intentionally radiates network traffic into space[7]. This makes it impossible to control who can receive the signals in any wireless LAN installation. In the wireless network, eavesdropping by the third parties is the most significant threat because the attacker can intercept the transmission over the air from a distance, away from the premise of the company.

5. ATTACKS ON WI-MAX AND COUNTERMEASURES

WiMAX (Worldwide Inter-operability for Microwave Access) network has the capability of working on many bands: 2.3 GHz, 2.5 GHz, etc, and provides scalability and mobility with high data rates with NLOS operation. It also provides strong security and strong QoS guaranteed services for data, voice, video, etc. However, in order for WiMAX to achieve a maturity level and become a successful technology, more research on security threats and solution to these threats need to be conducted. WiMAX has security vulnerabilities in both PHY and MAC layers, exposing to various classes of wireless attack including interception, fabrication, modification, and replay attacks. Some vulnerabilities of WiMAX originate from flaws of IEEE 802.16 on which WiMAX is based. A lot of problems and flaws have been fixed in the enhanced version but WiMAX still has some exposes.

In this section some possible threats or vulnerabilities will be reviewed and some solutions will be discussed.

1). Threats to the PHY layer

WiMax security is implemented in the security sub-layer which is above the PHY layer. Therefore the PHY is unsecured and it is not protected from attacks targeting at the inherent vulnerability of wireless links such as jamming, scrambling or water torture attack. WiMax supports mobility, thus it is more vulnerable to these attacks because the attackers do not need to reside in a fixed place and the monitoring solutions presented below will be more difficult.

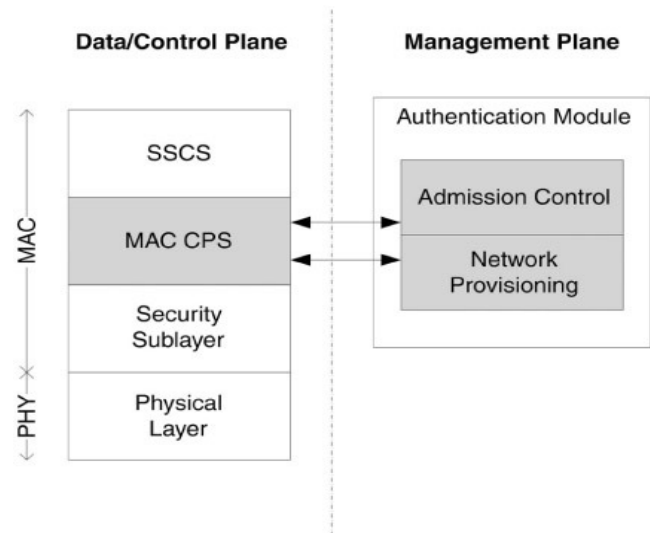


Fig. 1. Quality of Service Model of the IEEE 802.16

A. Jamming attack

Jamming can be described as an attack “achieved by introducing a source of noise strong enough to significantly reduce the capacity of the channel”. Jamming can be either intentional or unintentional. It is not difficult to perform a jamming attack because necessary information and equipments are easy to acquire and there is even a book by Poisel which teaches jamming techniques.

Solutions: We can prevent jamming attack by increasing the power of signals or by increasing the bandwidth of signals using spreading techniques such as frequency spread spectrum (FHSS) or direct sequence spread spectrum (DSS). Furthermore, since it is easy to detect jamming by using radio spectrum monitoring equipment and the sources of jamming are easy to be located by using radio direction finding tools, we can also ask help from law enforcement to stop the jammers.

B. Scrambling attack

Scrambling is a kind of jamming but only provoked for short intervals of time and targeted to specific WiMAX frames or parts of frames at the PHY layer. Attackers can selectively scramble control or management information in order to affect the normal operation of the network. Slots of data

traffic belonging to the targeted SSs can be scrambled selectively, forcing them to retransmit. It is more difficult to perform a scrambling attack than to perform a jamming attack due to “the need, by the attacker, to interpret control information and to send noise during specific intervals.

Solutions: Since scrambling is intermittent, it is more difficult to detect scrambling than jamming. Fortunately, we can use anomalies monitoring beyond performance norm (or criteria) to detect scrambling and scramblers.

C. Water torture attack

This is also a typical attack in which an attacker forces a SS to drain its battery or consume computing resources by sending a series of bogus frames. This kind of attack is considered even more destructive than a typical Denial-of-Service (DoS) attack since the SS which is a usually portable device is likely to have limited resources.

Solutions: To prevent this kind of attack, a sophisticated mechanism is necessary to discard bogus frames, thus avoiding running out of battery or computational resources.

D. Other threats:

In addition to threats from jamming, scrambling and water torture attacks, 802.16 is also vulnerable to other attacks such as forgery attacks in which an attacker with an adequate radio transmitter can write to a wireless channel. In mesh mode, 802.16 is also vulnerable to replay attacks in which an attacker resends valid frames that the attacker has intercepted in the middle of forwarding (relaying) process.

Solutions: WiMAX has fixed the security flaw of 802.16 by providing mutual authentication to defend these kinds of attacks.

2) Threats to the MAC layers

There are a lot of defects or flaws in WiMAX security solutions at the MAC layer. The vulnerabilities with MAC management messages are presented first in section 3.2.1 and section 3.2.2. Then vulnerabilities in authentication mechanism and some specific attacks are discussed.

A. Threats to Mac Management message in Initial network entry

The initial network entry procedure is very important since it is the first gate to establish a connection to Mobile WiMAX by performing several steps including: initial Ranging process, SS Basic Capability (SSBC) negotiation, PKM authentication and registration process.

The vulnerability of using Ranging Request-Response (RNG-REQ, RNG-RSP) messages: This message is used in

the initial ranging process. The RNG-REQ message is sent by a SS trying to join a network to propose a request for transmission timing, power, and frequency and burst profile information. Then, the BS responds by sending a RNG-RSP message to fine-tune the setting of transmission link. After that, the RNG-RSP can be used to change the uplink and downlink channel of the SS. There are several threats related to these messages. For instance, an attacker can intercept the RNG-REQ to change the most preferred burst profile of SS to the least effective one, thus downgrading the service. An attacker can also spoof or modify ranging messages to attack or interrupt regular network activities. This vulnerability can lead to a DoS attack which will be presented in details in 3.2.4 section of this report.

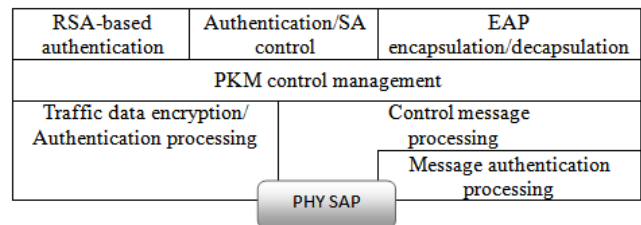


Fig. 2 MAC Security Sub layer

During the initial network entry process, many important physical parameters, performance factors, and security contexts between SS and BS, specifically the SBS negotiation parameters and PKM security contexts. Although the security schemes offered WiMAX include a message authentication scheme using HMAC/CMAC codes and traffic encryption scheme using AES based on PKMv2, these schemes are applied only to normal data traffic after initial network entry process. Subsequently, the parameters exchanged during this process are not securely protected, bringing a possible exposure to malicious users to attack.

Solution: a solution has been proposed to this vulnerability by using Diffie-Hellman key agreement scheme. In this approach, the Diffie-Hellman key agreement scheme will be used for SS and BS to generate a shared common key called “pre-TEK” separately and establish a secret communication channels in the initial ranging procedure. After that, the SBC security parameters and PKM security contexts can be exchanged securely.

B. Threats to Access network Security

In order to accommodate the requirements of WiMAX End-to-End Network Systems Architecture for mobile WiMAX network, the WiMAX forum defined network Reference Model (NRM) which consists of the following entities: Subscriber Station (SS), Access Service Network (ASN), and Connectivity Service Network (CSN)[8]. ASN consists of at least one BS and one ASN Gateway (ASN/GW) forming a complete set of network functions necessary to provide radio access to mobile subscribers. CSN consists of

AAA Proxy/Server, Policy, Billing, and Roaming Entities forming a set of network functions to provide IP connectivity services to subscribers.

Solutions: A countermeasure for this problem has been proposed by using a simple and efficient key exchange method based on Public Key Infrastructure (PKI). In this approach, all network devices have their certificates and a certificate chain for verification. The PKI structure is used as a method to obtain the correspondent's public keys and verify the certificates, thus enabling entities to create a shared secret key for establishing a secure connection.

C. Threats to Authentication

Many serious threats also arise from the WiMAX's authentication scheme in which masquerading and attacks on the authentication protocol of PKM are the most considerable.

Masquerading threat:

Masquerade attack is a type of attack in which one system assumes the identity of another. WiMax supports unilateral device level authentication which is a RSA/X.509 certificate based authentication. The certificate can be programmed in a device by the manufacturer. Therefore sniffing and spoofing can make a masquerade attack possible. Specifically, there are two techniques to perform this attack: identity theft and rogue BS attack.

- **Identity theft:**
An attacker reprograms a device with the hardware address of another device. The address can be stolen by interfering with the management messages.
- **Rogue BS attack:**
SS can be compromised by a forged BS which imitates a legitimate BS. The rogue BS makes the SSs believing that they are connected to the legitimate BS, thus it can intercept SSs' whole information. In IEEE 802.16 using PKMv1, the lack of mutual authentication prevents confirming the authentication of BS and makes Man-In-The-Middle (MITM) attack through rogue BS possible by sniffing Auth-related message from SS. However, it is difficult to successfully perform this kind of attack in WiMAX which supports mutual authentication by using PKMv2.

Attacks on the authentication protocols of basic PKM in 802.16 and its later version-PKMv2:

By adopting new version of PKM, WiMAX fixes many flaws in PKMv1 such as vulnerability to MITM due to the lack of mutual authentication. However, the newly proposed PKMv2 has been found to be also vulnerable to new attacks.

- **Attacks on basic PKM authentication protocol:** Attacker can intercept and save the messages sent by a legal SS and then perform a replay attack against the BS. The SS also might face with this kind of attack. In the worse case, since mutual authentication is not supported in basic PKM, BS is not authenticated. Therefore malicious BS can perform a MITM attack by making its own Auth-Reply message and gain the control of the communication of victim SS. Basic PKM has many flaws such that it provides almost no guarantees to SS about the AK. These problems have been fixed in the Intel Nonce version of PKM.
- **Attacks on Intel Nonce Version PKM:** In this version, nonce is a possible alternative to timestamp in authentication protocol. This approach does not protect a BS from a replay attack.
- **Attacks on PKMv2:**
This version provides a three-way authentication with a confirmation message from SS to BS. There are two possible attacks as follows. First, a replay attack can be performed if there is no signature by SS. Second, even with the signature from SS, an interleaving attack is still possible.

D. Other threats

Some serious attacks can exploit vulnerabilities in many aspect of the MAC layers. Two of the most destructive attacks can be MITM and DoS attacks.

Man in the middle attack:

Although WiMAX can prevent MITM attack through rogue BS by using PKMv2, it is still vulnerable to MITM attack. This possibility is due to the vulnerabilities in initial network entry procedure. It is known that WiMAX standard does not provide any security mechanism for the SSBC negotiation parameters. Through intercepting and capturing message in the SSBC negotiation procedure, an attacker can imitate a legitimate SS and send tampered SSBC response message to the BS while interrupting the communication between them. The spoof message would inform the BS that the SS only supports low security capabilities or has no security capability. If the BS still accepts, then the communication between the SS and the BS will not have a strong protection. Under these circumstances, the attacker is able to wiretap and tamper all the information transmitted. This kind of attack which they called "SINEP" can be dealt with a method based on Diffie-Hellman (DH) key exchange protocol.

Denial of Service attack:

Comprehensive surveys show that there are many vulnerabilities exposing IEEE 802.16e networks to DoS attacks such as unprotected network entry, unencrypted management communication, unprotected management

frame, weak key sharing mechanism in multicast and broadcast operations, and Reset-Command message).

Some of noticeable DoS attacks may include the following:

- *DoS attacks based on Ranging Request/Response (RNG-REQ/RNG-RSP) messages:*

An attacker can forge a RNG-RSP message to minimize the power level of SS to make SS hardly transmit to BS, thus triggering initial ranging procedure repeatedly. An attacker can also perform a water torture DoS by maximizing the power level of SS, effectively draining the SS's battery.

- *DoS attacks based on Mobile Neighbour Advertisement (MOB_NBR_ADV) message:*

MOB_NBR_ADV message is sent from serving BS to publicize the characteristics of neighbour base stations to SSs searching for possible handovers. This message is not authenticated. Thus it can be forged by an attacker in order to prevent the SSs from efficient handovers downgrading the performance or even denying the legitimate service.

- *DoS attacks based on Fast Power Control (FPC) message:*

FPC message is sent from BS to ask a SS to adjust its transmission power [9]. This is also one of the management messages which are not protected. An attacker can intercept and use FPC message to prevent a SS from correctly adjusting transmission power and communicating with the BS. He can also use this message to perform a water torture DoS attack to drain the SS's battery.

- *DoS attacks based on Authorization-invalid (Auth-invalid) message:*

The Auth-invalid is sent from a BS to a SS when AK shared between BS and SS expires or BS is unable to verify the HMAC/CMAC properly. This message is not protected by HMAC and it has PKM identifier equal to zero. Thus, it can be used as DoS tool to invalidate legitimate SS.

- *DoS attacks based on Reset Command (RES-CMD) message:*

This message is sent to request a SS to reinitialize its MAC state machine, allowing a BS to reset a non-responsive or malfunction SS. This message is protected by HMAC but is still potential to be used to perform a DoS attacks.

In order to prevent DoS attacks, we first need to fix the vulnerabilities in the initial network entry. It has been

suggested that the authentication mechanism should be extended to as many management frame as possible. They also suggest using digital signatures as an authentication method.

6. COMPARATIVE STUDY OF ATTACKS ON BLUETOOTH, WI-FI AND WIMAX

Table II summarizes the attacks on these three IEEE standards.

Bluetooth	Wi-Fi	Wi-Max
Bluesnarfing	Rogue Access Points	Jamming Attack
Bluebugging	Misconfiguration	Scrambling Attack
Bluejacking	Unmanaged Use of Wireless Technology Outside The Enterprise	Water Torture Attack
Denial of Service	Denial Of Service	Denial of Service
Car Whispering	Hackers	Forgery Attacks
Fuzzing Attacks	Spoofing And Session Hijacking	Initial Network Entry Threats
	Eavesdropping	Access Network Threats
		Authentication Threats

7. ACKNOWLEDGMENT

This work was supported by Department of Electronics and Telecommunication, Amity University, Uttar Pradesh.

REFERENCES

- [1] Gomes, Lee, "Many Wireless Networks Open to Attack," The Wall Street Journal Online, 27 April 2001
- [2] Lemos, Robert, "Wireless Networks Wide Open to Hackers," CNET News.com, 12 July 2001
- [3] Verton, Dan, "Flaws in Wireless Security Detailed," Computerworld, 16 July 2001
- [4] "The Bluetooth Blues", available at http://www.information-age.com/article/2001/may/the_bluetooth_blues
- [5] Phone pirates in seek and steal mission", Cambridge Evening News, available at: http://www.cambridge-news.co.uk/news/region_wide/2005/08/17/
- [6] Garcia, Andrew, "WEP Remains Vulnerable," eWEEK, 26 March 2001
- [7] Gomes, Lee, "Many Wireless Networks Open to Attack," The Wall Street Journal Online, 27 April 2001
- [8] "TLA Specification for Ranging and Authentication Process of IEEE Std 802.16-2004," <http://list.cs.northwestern.edu/802.16/>.
- [9] "A cost-based framework for analysis of denial of service in networks," Journal of Computer Security, vol. 9, no. 1-2, 2002.

Study on Combining 3G, Wi-Fi and Wi-max for Wireless Broadband

Nistha Rai¹, Kanak Priya², Neha Kumari³

^{1,2,3}M.Tech. (TSE), AITTM, Amity University, Noida

¹rai.nistha@gmail.com, ²kanak.priya90@gmail.com, ³nehakumari49@yahoo.co.in

Abstract: Many advanced data services are driving up wireless technology, which is further boosted by growth in advanced market segments. The wireless industry is evolving from a web of independent networks into a single integrated network with multiple standards; the expectation is that WI-FI, WI-MAX and 3G will coexist to enable a host of exciting new applications and business models. This paper focuses on concepts of these 3 technologies, their architecture and their comparison, which will help in boosting the wireless communication.

Keywords: wireless broadband, WIFI, WI-MAX, 3G)

1. INTRODUCTION

Wireless technology describes telecommunications in which electromagnetic waves, carry the signal over part or the entire communication path without cables. Wireless broadband refers to fixed wireless connectivity that can be utilized by enterprises, businesses, households and telecommuters who travel from one fixed location to another fixed location. Wireless broadband is an extension of the point-to-point, wireless-LAN bridging concept to deliver high-speed and high capacity pipe that can be used for voice, multi-media and Internet access services. However, there are many technologies available for providing broadband wireless access to the Internet, but the focus is on WI-MAX, 3G and WIFI due to their potential benefits. [5] This paper presents important features of WIMAX technology and an elaborated comparison of WIMAX with other contemporary technologies i.e. WI FI, 3G etc. Firstly, the potential features, advantages, disadvantages of WIFI, WIMAX and 3g are elaborated. Subsequently their architecture is discussed and finally WIFI and WIMAX are compared with each other along with their applications and conclusions are presented

2. II. OVERVIEW OF WI-FI, WI-MAX AND 3G

In this section, a brief overview of these technologies is discussed.

A. WI-FI

WI FI stands for wireless fidelity and generally refer to any type of 802.11 networks, whether 802.11b, 802.11a, 802.11g. WI-FI is a wireless technology that uses radio frequency to

transmit data through the air. W LAN access point, hub, or transmitter sends out a wireless signal that allows Wireless devices to access within a circle of roughly 100 meters. Zone around the transmitter is known as hot spot. Computers connected to WI FI receivers near a hot spot can connect to Internet at high speeds without cable. WI-FI refers to three types of wireless protocols that can work with each other: IEEE 802.11b ("Wireless B"), IEEE 802.11a ("Wireless A"), and the newer IEEE 802.11g ("Wireless G"). They can connect computers very fast: 11 Mbps for Wireless B, 54Mbps for Wireless A, and 54Mbps for Wireless G. [2] which are described in subsequent sections.

1) 802.11b

- * It is the longest, well-supported, stable, and cost effective standard, runs in the 2.4 GHz range that makes it prone to interference from other devices (microwave ovens, cordless phones, etc) and has security disadvantages
- * Limits to the number of access points to three.
- * It has 11 channels, with 3 non-overlapping, and supports rates from 1 to 11 Mbps.
- * It Uses direct-sequence spread-spectrum technology.

2) 802.11g

- * It is an extension of 802.11b, with the same disadvantages (Security and interference)
- * It has a shorter range than 802.11b
- * It is backwards compatible with 802.11b so it allows a Smooth transition from 11b to 11g
- * It is flexible because multiple channels can be combined for faster throughput, but limited to one access point
- * It runs at 54 Mbps,
- * Uses frequency division multiplexing technology

3) 802.11a

- * It is completely different from 11b and 11g.
- * It is flexible because multiple channels can be combined for faster throughput and more access points can be collocated
- * It has shorter range than 11b and 11g
- * It runs in the 5 GHz range, so having less interference from other devices
- * It has 12 channels, 8 non-overlapping, and supports rates from 6 to 54 Mbps

* It uses frequency division multiplexing technology. WI-FI is a trademark of the WI-FI Alliance (formerly the Wireless Ethernet Compatibility Alliance), the trade organization that tests and certifies equipment compliance with the 802.11 standards.

Advantages of Wi-Fi

- * It Uses an unlicensed portion of the broadcast spectrum, and requires less regulatory controls in many countries.
- * It frees network devices from cables, allows a more dynamic network to be grown.
- * Many reliable and bug-free WI-FI products are available in the market.
- * Competition amongst vendors has lowered prices considerably since their inception. It is possible to move without breaking the network connection while connected on a WI-FI network.

Disadvantages of WI-FI

- * The 802.11g and 802.11b standards of WI-FI use the 2.4 GHz spectrum, which is crowded with other devices such as blue tooth, microwave ovens, and cordless phones. It may cause degradation in performance. Other devices, which use microwave frequencies such as certain types of cell phones, can also cause degradation in performance.
- * Power consumption is high compared to other standards, making battery life and heat a matter of concern. Sometimes Users cannot configure it properly. WI-FI commonly uses WEP (wired equivalent privacy) protocol for protection, which can be easily breakable even when properly configured. Newer wireless solutions are providing support for the superior WPA (WI FI protected access) protocol (implementation of the 802.11i protocol), though many systems still employ WEP. By adopting 802.11i protocol makes available a better security scheme for future use when properly configured. WI-FI networks have limited range. A typical WI-FI home router using 802.11b or 802.11g have a range of 150 ft (46 m) indoor and 300 ft (92 m) outdoors. [5]

B. WI-MAX

Wi-MAX, an acronym that stands for Worldwide Interoperability for Microwave Access and is based on point to point broadband wireless access and working on the group no. 16 of IEEE 802i.e. IEEE 802.16

1) Features of Wi-Max

Uses Microwaves for The Wireless Transfer Of Data

- * It stands for Wireless (WI) microwave access (MAX)
- * It is used for high-speed, wireless networking at distances of a few kilometres

- * It Uses OFDM (which allows for non line-of-sight communications and addresses multipath issues)
- * It Includes TDD and FDD duplexing support
- * It has flexible channel sizes (3.5MHz, 5 MHz, and 10 MHz) [3]

2) Wi-Max Promises

- * Up to a ten (10) mile range without wires
- * Broadband speeds without cable or TI
- * Handles "last mile" access in remote areas
- * Licensing and equipment due in 2005
- * Affordable technology

3) Standard based.

The Wi-MAX standard has fragmented into two variants: 802.16a the original Wi-MAX standard, which can transfer at up to 70Mbps over distances of as much as 30 miles using the 10GHz and 66GHz spectrums. And 802.16e A more recent development which will operate in the 2GHz - 6GHz licensed bands, bringing the possibility of mobile devices using the technology.

4) Frequency under considerations

Three Frequencies under consideration for WI max 5.8GHz unlicensed (same as for Wi-Fi). However, the fact that it is unlicensed limits the transmission distance 2.5GHz and 3.5GHz have been targeted for licensed use throughout the world. Wi-Max Regulatory Task Force coordinates with Radio

Regulators across the World and Wi-Max representation at International Bodies - Such as ITU, WRC, CEPT, ERO and FCC

C. 3G (3rd generation)

3G stand for 3rd generation mobile telephone systems. It is a technology for mobile service providers. 3G combines high speed mobile access with Internet Protocol (IP) based services. 3G can use a variety of present and future wireless network technologies.

Evolution of 3G

The first mobile services were analog. Mobile services began to emerge in the 1940s, the first mass market mobile services in the U.S. were based on the AMPS (Advanced Mobile Phone Service) technology. IT is referred as first generation wireless. The FCC licensed two operators in each market to offer AMPS service in the 800-900MHz bands. In the 1990s, mobile services based on digital mobile technologies were known as second generation (2G) of wireless services. In the U.S., these were referred as

Personal Communication Systems (PCS) and the technologies used were TDMA (Time Division Multiple Access), CDMA (Code Division Multiple Access) and GSM (Global System for Mobile Communications). [5] From 1995 to 1997, the FCC auctioned off PCS spectrum licenses in the 1850 to 1990 MHz band. CDMA and TDMA were deployed in the various parts of the U.S., while GSM was deployed as the common standard in Europe. The next or Third Generation (3G) mobile technologies hopes to support higher bandwidth digital communications and are expected to be based on one of the several standards included under the ITU's IMT-2000 umbrella of 3G standards.[6]. In the next section, the architecture of various technologies are described.

3. ARCHITECTURE

A. WI-FI

1) WI-FI building block.

Block diagram shows mapping of the IEEE 802.11 requirements into a functional Wi-Fi building block. The Wi-Fi building blocks are:

1. Antenna
2. Access Point (AP)
3. Router
4. Internet

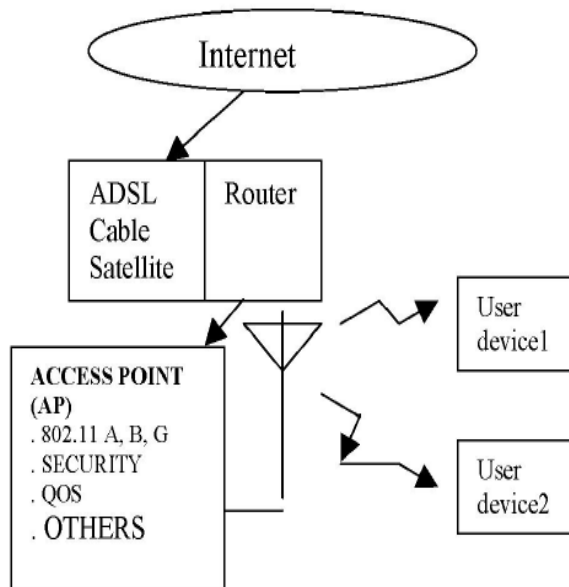


Fig 1: WI FI Sketch

2) Wi-Fi Network Cell

Wi-Fi network cell is of 802.11b standard. This standard defines 11 channels. The RF reach of each channel is about

160 ft in doors or about 300ft outdoors. There are three non-overlapping channels (channels, 1, 6, and 11). Channels 1, 6, and 11 are typically used to cover a large area.

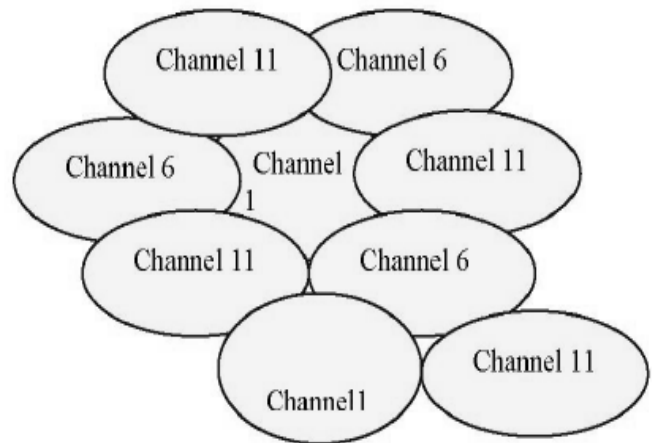


Fig 2: channel allocation for WI FI

3) Network architecture

A network is established when a station(s) and APs have recognized each other and established a communication link. A network can be configured in two basic ways:

- * Ad hoc (Peer-to-peer) network
- Infrastructure network

Ad hoc (Peer-to-peer) network

In this configuration, two or more stations can talk to each other without an AP. This arrangement is referred to as an Independent Basic Service Set (IBSS). Access to the wired network (Internet) is accomplished at the station that has the Internet access port.

Infrastructure network

This configuration consists of multiple stations connected to an AP. The AP acts as a bridge to the wired network. This arrangement is referred to as a Basic Service Set (BSS).

Wi-Fi supports various network types:

- * Hub and spoke
- * Mesh network

B. WIMAX

Wi MAX is based on point-to-point broadband wireless access and working on the group no. 16 of IEEE 802i.e. IEEE 802.16 [3] WiMAX is known as one of the broadband fixed wireless access solutions for the "Last mile" and designed to address the MAN market.

1) WiMAX network architecture

WiMAX architecture is similar to the cellular telephony in that a service area is divided into cells. WiMAX is able to operate in a Line Of Sight (LOS) and near or non LOS (NLOS) access approach.

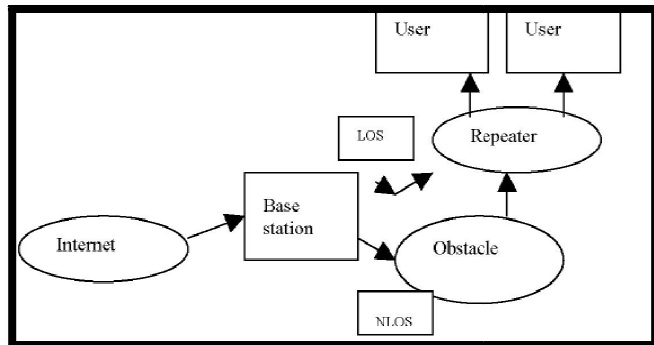


Fig3- WiMAX sketch

Wi Max Potential Features

OFDM (Orthogonal Frequency Division Multiplexing) OFDM is a technique, which is based on multi carrier modulation (MCM) and Frequency Division Multiplexing (FDM). OFDM is a modulation or multiplexing method. In multi carrier, (fig 4) modulation signal bandwidth is divided into parallel subcarriers or narrow strips of bandwidth. Here subcarriers (fig 5) are overlapping, i.e. OFDM uses subcarriers that are mathematically orthogonal, and information is sent on Parallel overlapping subcarriers, from which information can be extracted individually. This property reduces interference caused by adjacent carriers

Fig 4 Multi carrier

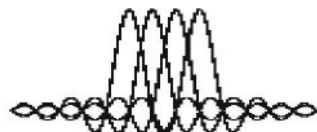


Fig 5 Single carrier



* Sub Channelization:

Adaptive Modulation and Coding (AMC) is in a multipath environment that gives OFDMA advantages as the flexibility to change the modulation for specific. Sub-channels allow optimizing at the frequency level. Another alternative would be to assign those sub channels to a different user who may have better channel conditions for that particular sub-channel. This allows users to concentrate transmitted power on Specific sub channels, resulting in improvements to the uplink budget and providing greater range. This Technique is known as Space Division Multiple Access (SDMA). Fig(6) shows selection of sub channels depending on the

received signal strength. Those Sub-channels on which the user experiences fading are avoided and power is concentrated on channels with better channel conditions. The signals on the top indicate the received signal strength, while the bottom indicates which sub-carriers are then chosen for each Signal. With OFDMA, the client device chooses sub channels. Based on geographical locations with the potential of eliminating the impact of deep fades.[4]

4. COMPARISONS

A. Differences between WI-FI and WI-MAX

- * WI-FI channels occupy a fixed width of the spectrum WI-Max will allow channel sizes to be decided based on requirements
- * WIMAX works in the licensed spectrum (10-66 GHz with some systems operating in bands between 2 GHz and 11 GHz). Wi-Fi works in the unlicensed ISM band.
- * WI-FI has a range of about 100m. Wi-Max will provide services over several kms. WI-MAX is designed to provide a higher level of reliability and quality of service than Wi-Fi
- * Wi-MAX is a MAN protocol that will be a wireless alternative to DSL and TI level services for last mile broadband access it will act as a backbone technology for 802.11 hotspots.
- * WI-MAX is Easier to install than WI-FI (no need for careful alignment of access points) - goal of easy self-installation various standards FOR WI-FI, WI-MAX.

Table 1: WIFI AND WIMAX

Standard	Max throughput	Freq. Bands	Application Area
802.11 a	54 Mbps	5.5 GHZ	LAN (Fixed & Mobile)-100 m
802.11 b	11 Mbps	2.4 GHZ	LAN (Fixed - Mobile) – 100 m
802.11g	54 Mbps	2.4 GHZ	LAN (Fixed-Mobile) –100 m
802.16 (Wi max)	70 Mbps	5 GHZ, 11 GHZ	WAN (Fixed) – 50 Km

Emerging WI-FI broadband applications

1) The Benefits of Wi-Fi / WiMAX in rural regions: WI-FI deployment in rural communities offers opportunity on many fronts. Wi-Fi can cost effectively be deployed in weeks. Visitors and residents can travel around the community and maintain Internet access through their handheld wireless devices. One advantage Wi-Fi has is the inherent enormous bandwidth access when compare to the legacy Internet accesses. Wi-Fi communities are taking advantage of this vast capacity to build their own Wi-Fi broadband private virtual network

2) Web-based events: Web-based hosting events offer several possibilities for communities to merchandize their goods and services. Such community sites could attract new customers and potential visitors can familiarize themselves with community interests, talents, art, and business potentials.

3) Video chat: Video chat is a powerful tool for increasing interaction between users in a public setting. It encourages participants to interact in a friendly social setting. In a wireless community, town hall meetings could enjoy high 'virtual' attendance.

4) Wi-Fi Killer application: Voice over Wi-Fi is an emerging application that has been referred to as the "killer application".

B. WIFI VS 3G

1) Similarities

Both are wireless: Both technologies are wireless such that

(1) Avoids need to install cable drops to each device when compared to wire line alternatives;

(2) It facilitates mobility.

(i) Both are access technologies

Both 3G and Wi-Fi are access or edge-network technologies. They offer alternatives to the last-mile wire line network. Beyond the last-mile, both rely on similar network connections and transmission support infrastructure.

(ii) Both offer broadband data service

Both 3G and WIFI support broadband data service, the data rate offered by Wi-Fi (11Mbps) is substantially higher than the couple of 100 Kbps expected from 3G services.

2) DIFFERENCES

In this section, we consider several of the important ways in which the Wi-Fi and 3G approaches to offering broadband wireless access services are substantively different.

(i) Current business models deployment is different.

3G represent an extension of the mobile service provider model, whereas Wi-Fi comes out of the data communications industry (LANs), which is a by-product of the computer industry

(ii) Spectrum policy and management.

One of the key distinctions between 3G and Wi-Fi is that 3G and other mobile technologies use licensed spectrum, while Wi-Fi uses unlicensed shared spectrum. This has important implications for

(i) Cost of Service

(ii) Quality of Service (QoS)

(iii) Congestion Management,

(iv) Industry structure

(iii) Status of technology development is different.

The two technologies differ with respect to their stage of development in a number of ways as-

1. Deployment Status

While 3G licenses have been awarded in a number of markets at a high cost, there is only limited progress with respect to service Deployment. Large base of Wi-Fi networking equipment has installed that is growing rapidly as Wi-Fi vendors have geared up to push wireless home networks using the technology and are unlicensed.

2. Embedded Support for Services

Difference between 3G and Wi-Fi is their embedded support for voice services. 3G is expressly designed as an upgrade technology for wireless voice telephony networks, so voice services are an intrinsic part of 3G. In contrast, Wi-Fi Provides a lower layer data communications service. For example, with IP running over Wi-Fi it is possible to support Voice-over-IP telephony. Another potential advantage of 3G over Wi-Fi is that 3G offers had better support for Secure/private communications than does Wi-Fi.

3. Standardization

Broadly, formal standards picture for 3G is clearer than for WLAN. For 3G, there is a relatively small family of internationally sanctioned standards collectively referred to as WCDMA. However, there is still uncertainty as to which of these (or even if multiple ones) will be selected by service providers. In contrast, Wi-Fi is one of the family of continuously evolving 802.11x Wireless Ethernet standards, which is itself one of many WLAN technologies that are under development

4. Service/Business Model

3G is more developed than Wi-Fi as a business and service model. It represents an Extension of the existing service provider industry to new services. Whereas, Wi-Fi is more developed with respect to the upstream supplier markets (with respect to WLAN equipment) and is commoditized. Communicating with a 3G base station at a long distance but with reduced bandwidth or Communicating with a Wi-Fi

base station at a short distance but at a much higher data rate will both consume batteries at a faster rate.

TABLE 2: 3G AND WIFI

	3g	WI FI
Standard	WCDMA, CDMA 2000	IEEE 802.11
Max speed	2mbps	54mbps
Operations	Cell phone companies	Individuals
License	Yes	No
Coverage area	Several km	About 100m
Advantages	Range, mobility	Speed, cheap
Disadvantage	Slow, expensive	Short range

C. WIMAX VS 3G

The differences between WIMAX and 3G are shown in the table.

TABLE 2: 3 G AND WIMAX

	3g	WI MAX
Standard	WCDMA, CDMA2000	IEEE 802.16
Max speed	2mbps	10 to 100mbps
Operations	Cell phone companies	Individuals
License	Yes	Yes/no
Coverage area	Several km	Several km
Advantages	Range, mobility	Speed, long range
Disadvantage	Slow, expensive	Interference

5. CONCLUSION

Wi-Fi is good for competition and has an extension to WiMAX. Wi-Fi and 3G complement each other for a mobile provider. Success of Wi-Fi is potentially good for multimedia content. At a long distance but with reduced bandwidth one can communicate with 3G efficiently. On the other hand, Wi-Fi can be used to communicate at a short distance but at a much higher data rate. WiMAX is used to cover large area, which overcome the disadvantage of Wi-Fi as discussed in the paper. Hence, depending on the requirements, the technologies can coexist.

6. ACKNOWLEDGMENTS

It is a great pleasure to acknowledge my profound sense of gratitude to my project guide Mrs. Neha Arora, Assistant Professor, AITTM AUUP for her valuable and inspiring guidance, comments, suggestions and encouragement throughout this paper.

REFERENCES

- [1] Al Prendergast Why Wi-Fi? LV-CCLD August 12, 2004
- [2] RFM (RF Monolithics, inc) WiMAX (Worldwide Interoperability for Microwave Access)
- [3] Understanding WiMAX and 3G for Portable/Mobile Broadband Wireless Technical White Paper A Technical Overview and Comparison of WiMAX and 3G Technologies December 2004
- [4] William Lehr Lee and W. McKnight Wireless Internet Access: 3G vs. Wi-Fi? August 23, 2002
- [5] IMT -2000, GENEVA 2001-2002 Bandwidth

A Comparison Study of Dynamic Addressing Techniques in Mobile Ad hoc Networks

Navin Paudel

Amity Institute of Telecom and Management
Amity University, Uttar Pradesh (AUUP) Campus, Sector-125
Noida-201303 (U.P), India.
paudelnavin@gmail.com

Abstract: Dynamic address assignment enables nodes in mobile ad hoc networks to obtain a routable address without the need for any explicit configuration. It provides a means for nodes to communicate without any centralized infrastructure, and provides a mechanism for dynamic network membership. Therefore, a protocol is needed to perform the network configuration automatically and in a dynamic way, which will use all nodes in the network (or part thereof) as if they were servers that manage IP addresses. This article reviews the major proposed auto-configuration protocols for mobile *ad hoc* networks. This work also includes a comparison of auto-configuration protocols for mobile *ad hoc* networks by specifying the most relevant metrics, such as a guarantee of uniqueness, overhead, latency, dependency on the routing protocol and uniformity.

1. INTRODUCTION

A *Mobile Ad hoc NETWORK* (MANET) is a set of mobile nodes which communicate between themselves through wireless links. In contrast with conventional networks, a MANET does not need any previous infrastructure, since nodes rely on each other to operate themselves, forming what is called multi-hop communication. Such networks have more problems and disadvantages than a conventional network. The topology of mobile networks may change quickly and in an unpredictable way. To communicate with each other [1] the *ad hoc* nodes need to configure their interfaces with local addresses which are valid within the *ad hoc* network. The *ad hoc* nodes may also need to set global routing addresses to communicate with other devices on the Internet. From the perspective of the IP layer, an *ad hoc* network presents itself as a multi-hop level 3 network constituted by a collection of links.

2. AUTO-CONFIGURATION IN MOBILE AD HOC NETWORK

The nodes of a network need some mechanism to interchange messages with each other. The TCP/IP protocol allows the different nodes from the network to communicate by associating a distinct IP address to each node of the same network. Mobile *ad hoc* networks do not have such a centralized entity able to carry out this function. Therefore,

some protocol that performs the network configuration in a dynamic and automatic way is necessary, which will utilize all the nodes of the network (or only part of them) as if they were servers which manage IP addresses. Due to the dynamic topology of mobile *ad hoc* networks auto-configuration protocols are faced with various problems in guaranteeing the uniqueness of IP addresses and in allowing network partitioning and merging. To guarantee the correct functioning of the network, the protocols strive to achieve the following objectives [1-3]:

- *Assign unique IP addresses:* Ensure that two or more nodes do not obtain the same IP address.
- *Function correctly:* An IP address is only associated with a node for the time that it is kept in the network. When a node leaves the network, its IP address should then become available for association to another node.
- *Fix the problems derived from the loss of messages:* In case of any node failure or if message loss occurs, the protocol should operate quick enough to prevent two or more nodes from having the same IP address.
- *Allow multi-hop routing:* A node will not be configured with an IP address if there aren't any available in the whole network. Thus, if any node of the network has a free IP address, it has to associate itself with the node which is requesting an IP address, even though it is at two-hops of distance or more.
- *Minimize the additional packet traffic in the network:* The protocol must minimize the number of packets exchanged among the nodes in the auto-configuration process. In other words, control packets traffic must cause as little harm as possible to the data packet traffic, given that in the extreme case, the network performance would decrease.
- *Verify the existence of competing petitions for an IP address:* When two nodes request an IP address at the same time, the protocol must carry out the pertinent treatment so that the same IP address is not given to two nodes.

- *Be flexible to partitioning and merging of the mobile ad hoc network:* The protocol must be able to achieve the union of two different mobile *ad hoc* networks as well as the possible partitioning into two networks.
- *Conduct synchronization:* The protocol must adapt itself to the rapid changes of the wireless network topology due to the frequent mobility of the nodes. The synchronization is carried out periodically to ensure the configuration of the network is as up to date as possible.

3. CLASSIFICATION OF AUTO-CONFIGURATION PROTOCOLS

The auto-configuration protocols may be classified according to address management:

- *Stateful:* The nodes know the network state, *i.e.*, they keep tables with the IP addresses of the nodes.
- *Stateless:* The IP address of a node is managed by itself. Generally they create a random address and perform a process of duplicated address detection steps to verify their uniqueness.
- *Hybrid Protocols:* They mix mechanisms from the previous ones to improve the scalability and reliability of the auto-configuration. Their algorithms have a high level of complexity.

3.1. Stateful Protocols

3.1.1. MANETConf

MANETConf [4], which is an improvement of [5], is based on existence of a common distributed table so that all the nodes are able to assign IP addresses. When a node wants to join the network, it sends broadcast messages to other nodes and the first one which replies to the message, chooses it as an initiator node and it can supply an IP address. The initiator node chooses one of the free IP addresses located in the network

3.1.2. DAAP

Dynamic Address Allocation Protocol (DAAP) [6] is based on the concept of address assignment by a leader. The leader functionality is shared among all network nodes. When a new node joins the network, it becomes the leader until the next node joins. The leader maintains the highest IP address within the *ad hoc* network and a unique identifier is associated with the network. Each node stores the highest IP address, which is that of the leader, and periodically sends HELLO messages to its neighbors. These HELLO messages include the network identifier so that any merging and partitioning can be detected. When a node receives a HELLO message with a different network ID, merging is detected, if a node does not receive the message that

contains the current network ID, then after a timeout, a partition is detected.

3.1.3. Moshin and Prakash's Protocol

The Moshin and Prakash proactive scheme [7] tries to fix the problem of IP address assignment by binary division of free address blocks. The address assignment process can come from any node. Each node which initiates the network generates a random number called *PartitionID* which will be a network identification number. In this manner, when a partitioning or merging of the network occurs, the first node which leaves from original network will create another *PartitionID*. At the time the two networks with different *PartitionID* are joined; firstly the consistency of its IP address will be checked. This process consists of verifying the existence of two nodes with the same address. In case affirmative, a change of the node belonging to the network with less free IP address range is produced. The new IP address will belong to the highest network range with the biggest free address number. The major drawback of this protocol is that the synchronization depends on the existence of a reliable *broadcast* and such a thing does not exist in a distributed mobile environment, thus one can question the robustness of this protocol.

3.1.4. Thoppian and Prakash's Protocol

An improvement of the previous scheme (particularly in terms of synchronization) can be found in [8], where Thoppian and Prakash propose a dynamic address assignment based on a so-called *buddy system* that manages mobility of nodes during address assignment, message loss, network partitioning and merging. However, the IP address allocation can generate a high overhead of control messages while it does a global search and the address recovery (to avoid missing addresses) requires diffusion messages by a *flooding* process. In addition, union and partition may incur in high overhead because of the global nature of this protocol.

3.1.5. EMAP

Extensible Manet Auto-configuration Protocol (EMAP) [9] is an auto-configuration protocol based on the idea of a protocol of REQUEST/REPLAY messages. The main advantage of this protocol is the possibility of doing it extensible, *i.e.*, it can include new functionalities in the future that are analyzed in a theoretical way, such as *Domain Name Server* (DNS). This protocol also considers the possibility of exterior communications to the mobile *ad hoc* network via Internet. The route discovery mechanism among nodes is similar to the *Ad Hoc On-Demand Distance Vector* (AODV) [10] protocol. When a node wishes join the network, it randomly generates two valid IP addresses (with network known addresses), considering them as temporary and tentative addresses. These IP addresses are encapsulated

into a *Detection Address Detection REsPonse* (DAD_REP) message to know whether it is a valid address. The node keeps waiting for a *Detection Address Detection REQuest* (DAD_REQ) message. If time runs out for this message, the node assumes that it can use its tentative address as unique, and assigns it to its network interface. If this node receives a DAD_REP message to its temporary address and this message contains the source with the tentative address that had been proposed, the node knows that this tentative address is being used and the previous process begins creating another pair of addresses again.

3.2. Stateless Protocols

3.2.1. Process of Duplicated Address Detections

Duplicate Address Detection (DAD) is a process which uses the protocols to check the uniqueness of IP addresses. This process takes a relatively long time to complete, so several solutions have been implemented to reduce it. There are three kinds of DAD processes:

- *Strong Duplicate Address Detection (SDAD)* [11]: Is the base of the Stateless protocols. It consists of a simple mechanism whereby the node chooses two IP addresses: temporary and tentative. It will only use the temporary address for the initialization while it detects if the tentative one is unique or not. The detection method consists of sending a message ICMP destined directly to this address. If it receives a response, this IP address is being used so the process will be resumed. If it does not receive a response, the message will be sent a certain number of times to make sure that the address is unique. By being a very simple mechanism, it does not ensure the uniqueness of the IP address since the process limits itself to only them phase of initialization, and it would not work for temporary disconnections or losing of the network. Moreover, when the network is long and only a few free IP addresses remain, it increases the overhead until it finds a unique IP address.
- *Weak Duplicate Address Detection (WDAD)* [12]: It establishes the idea of tolerating the duplicated address in the network for a period of time. For that, every node when it is being initiated itself will create a key that it will always send along with its IP address. When a node receives a message, it will check whether this IP address is already assigned in its table and will look whether the keys coincide, if they do not coincide, it will mark that address as invalid and actions will be taken so that they are unique (these actions are not defined in WDAD).
- *Passive Duplicate Address Detection (PDAD)* [13]: The idea is based on sending control information instead of detecting or solving duplicated IP addresses, every node investigates and deduces whether a duplicated address exists by events that would never happen if all the IP

addresses were unique. Three passive detections are proposed, which are necessary for correct functioning of the detection:

- *Sequence Numbers (PDAD-SN)*: This system is based on the idea that the routing protocols use sequence numbers in its messages to update the routes. Using these sequence numbers, and the idea that two nodes with a distance between them of two-hops do not have the same neighborhood, some conflicts are solved. Also, it has taken into consideration the possibility that these sequence numbers reach the maximum and numbers start from zero again.
- *Locality Principle (PDAD-LP)*: Lower power than the previous one, is based on the frequency of updating the route tables. On the basis of this frequency it can detect m duplicate addresses by taking a time threshold to display the status of the route tables. It's necessary to take into account the protocol routing used. We must consider different thresholds, since it can be the case that two messages with the same source are confused with a duplicate address, if the time is too short and, therefore, the protocol modifies the routes too fast.
- *Neighborhood (PDAD-NH)*: Taking into account that a node knows its neighbours, and they have sent a package of the state of its link, it differs if there is conflict or not depending on whether it is in a package, the source of this message is a neighbour, and contains the address. The advantage is that it does not add overhead to network, but it can only be used with proactive routing protocols.

3.2.2. APAC

Agent based Passive Auto-configuration (APAC) [14] is an auto-configuration protocol based on PDAD. Its main feature is the use of certain nodes which centralize the distribution of addresses. The mechanism by which a node configures its IP address upon entering the network consists of asking if it has some node type *Address Agent* (AA) within a one hop distance. In that case, the AA node will give it an IP address. If it does not receive a response from any AA node, the incoming node is configured to operate in AA mode, and it will be a server of addresses for the next incoming nodes. When it is configured as AA, the node randomly generates an identifier number *agentID*, in order to form a table with the addresses that it will assign to the nodes that arrive. These addresses have the form *agentID* + *hostID*. When a node is moved in the network and leaves the radio coverage of AA proportionately to its IP address, this node must ask for another address from another node AA within its new radio coverage. The detection of duplicated addresses is undertaken in the PDAD process. Once any conflict has been detected, the AA which assigned the conflictive address is informed. This AA node will then generate a new *agentID* and it will warn all its dependant

nodes to change their address from *agentID* + *hostID* type to the new *agentID*. In case of partitioning, the nodes AA will mark addresses which have left the network as free. In the case of the merging of two networks, the mechanism for the detection of duplicate addresses continues working properly.

3.2.3. AROD

In *Address auto-configuration with address Reservation and Optimistic duplicated address Detection* (AROD) [15], the address reservation is based on the existence of nodes that have an IP address reserved to deliver it to the new nodes that enter. Two types of nodes will exist:

- Agents type 1 with a reserved IP address, apart from the IP address that has its network interfaces. When a node joins the network, this reserved IP will be assigned to it immediately.
- Agents type 2, which do not have reserved IP addresses. If a node that joins newly asks one of these for an IP address, this node borrows the reserved address of one of its neighbors who is of type 1, and it is assigned to the new one immediately. The IP addresses are unique; therefore, the two nodes turn into node type 1.
- If only one is unique, the one that gave the IP address will be turned into node type 1.
- If none is unique, the two nodes will remain as type 2.

This protocol considers its process of duplicated address detection as an optimistic DAD process, because it is only carried out once when a node joins and a reserved IP address is assigned to it.

3.2.4. AIPAC

Automatic IP Address Configuration in Mobile Ad Hoc Networks (AIPAC) [16] is a protocol for IP address auto-configuration using a reactive approach in the IP address assignment. Each network is identified with its *NetID*. When two networks merge, and the merger is persistent, the *NetID* should be unified. To accomplish that, it uses a gradual fusion mechanism. This allows a node to pass from a *NetID* to another one, according to network changes observed by the node. This procedure allows a homogeneous system to be made in the case of multiple overlapping networks, according to the evolution of their topologies. This protocol does not guarantee the uniqueness of the assigned IP addresses, but it ensures that messages are routed correctly. Each node in AIPAC is aware of its neighbouring radio, so the amount of information stored by the node is limited to the nodes within the radios distance.

3.3. Hybrid Protocols

3.3.1. HCQA

Hybrid Centralized Query-based Autoconfiguration (HCQA) [17] was the first hybrid auto-configuration

protocol. A node that wants to join the network undergoes a SDAD process. If the process is successful, the node will have to register its tentative IP address with an *Address Authority*. When the network is created, the first node becomes an Address Authority, it chooses a unique identifier for the network (e.g., MAC address) and advertises it periodically by *broadcast* messages to identify the network. If a node does not receive it, it is assumed that the network has been divided and it will create its own network becoming Address Authority. This protocol adds robustness to the SDAD process, it ensures no duplicity of IP addresses and it also provides a good mechanism for network *partitioning*. However, it has two main problems, firstly the overhead produced by the SDAD process and periodic messages of Address Authority, and secondly, the network depends on a central entity with which all nodes must communicate directly in order to register its IP address, so that much latency is added at the joining of nodes to the network.

3.3.2. PACMAN

Passive Autoconfiguration for Mobile ad hoc Networks (PACMAN) [18] is a passive auto-configuration protocol for MANET. It uses elements from stateless and stateful protocols, so it could be considered somewhat hybrid. Its operation is based on each node assigning itself an address when joining the network, and passive monitoring of communications for the duplicate address detection. To achieve the minimum overhead in the communications, the information is shared among different network layers. Specifically, the information handled by the routing protocol is monitored. The method used to choose the own IP address consists of a probabilistic algorithm. The probability of attempting to choose an IP address currently in use by another node is close to zero, this algorithm takes into account, among other factors, an assignment table. This table is created with information from the routing protocol on IP addresses already in use. PACMAN uses the PDAD process to monitor the communications in search of duplicated addresses. This is necessary because the mechanism used for address assignment does not guarantee uniqueness (even if it attempts to reduce the probability of collision), and it may cause merging of networks containing nodes with the same IP address. Broadly speaking there are two types of events that indicate duplicity of IP addresses: firstly, we have the events that never occur if the address is unique, and always occur if the address is duplicated. These events confirm that there is a problem detected. On the other hand, we have the events that occur rarely if the address is unique, and often if the address is duplicated. In this way the possibility of problems is detected, so it is probabilistic algorithms. When it detects that two nodes are using the same IP address, it reports the problem to one of them using a *unicast* message to change its address. Moreover, it takes into account the problem of changing an address that has some

communication going on. To fix this, when changing an address, a node notifies the nodes with which it has ongoing communications of its new IP address, so that they can make an encapsulation of the messages properly.

4. PERFORMANCE EVALUATION OF AUTO-CONFIGURATION PROTOCOLS

4.1. Performance Metrics for the Evaluation of Auto-Configuration Protocols

Zhou *et al.* [19] define some parameters that can be used to analyze the performance of an auto-configuration protocol for *ad hoc* networks (see Table 1).

4.2. Performance Evaluation of Studied Protocols

The submitted protocols share some common characteristics. However, they also differ in a wide range of issues. Table 2 presents a comparison of the characteristics of IPv4 addressing protocols.

Table 1. Evaluation metrics.

Metrics	Description
Uniqueness	Each MANET node must have a unique IP address for each network interface because duplicate addresses can cause serious routing problems.
Overhead	Exchanged packet number to obtain an IP address.
Latency	Node timeout to obtain the IP address.
Routing Protocol Independence	Auto-configuration protocols can work in two ways: leaning on a routing protocol to allow the routing of the new nodes joining the network, or regardless of routing algorithm.
Uniformity	All nodes perform the same function in the auto-configuration process.

Table 2. Comparison of the characteristics of IPv4 addressing protocols.

	Protocol	Guarantee of Uniqueness	Overhead	Latency	Dependent Routing	Uniform
Stateful	ManetConf	No	High	High	No	Yes
	DAAP	Yes	Medium	Medium	No	No
	Buddy System	Yes	Medium	Medium	No	Yes
	EMAP	No	Low	High	No	Yes
Stateless	SDAD	No	High	High	No	Yes
	WDAD	No	Medium	Low	No	No
	PDAD	No	High	High	No	Yes
	APAC	No	High	High	Yes	No
	AROD	Yes	High	High	No	No
	AIPAC	No	High	High	No	No
Hybrid	HCQA	Yes	High	High	Yes	No
	PACMAN	No	High	High	Yes	Yes

5. CONCLUSION

The nodes of a network need a mechanism to exchange messages. The TCP/IP protocols can allow the different nodes of the same network to be associated with a different IP address. Due to the dynamic topology of mobile *ad hoc* networks (constant movement of the nodes that can enter and leave the network frequently or even simultaneously), auto-configuration protocols face many problems with guaranteeing the uniqueness of IP addresses. I studied current solutions by categorizing and qualitatively analyzing

scalability and other performance properties of the approaches.

REFERENCES

- [1] Bernardos C, Calderon M, Moustafa H. Ad-Hoc *IP Autoconfiguration Solution Space Analysis*. Nov, 2008. Internet Draft; Available online: <http://tools.ietf.org/pdf/draft-bernardos-autoconf-solution-space-02.pdf> (accessed on 25 November 2010).
- [2] Bernardos C, Calderon M, Moustafa H. Survey of IP Address Autoconfiguration Mechanisms for MANETs.

- Nov, 2008. Internet Draft; Available online: <http://tools.ietf.org/html/draft-bernardos-manet-autoconf-survey-04> (accessed on 25 November 2010).
- [3] Bernardos C, Calderon M, Moustafa H. Evaluation Considerations for IP Autoconfiguration Mechanisms in MANETs. Nov, 2008. Internet Draft; Available online: <http://www.it.uc3m.es/cjbc/papers/draft-bernardos-autoconf-evaluation-considerations-03.txt> (accessed on 25 November 2010).
- [4] Nesargi S, Prakash R. MANETconf: Configuration of Hosts in a Mobile *Ad Hoc* Network. Proceedings of IEEE INFOCOM 2002; New York, NY, USA. June 2002; pp. 1059–1068. Available online: <http://www.utdallas.edu/~ravip/papers/infocom2002.pdf> (accessed on 25 November 2010).
- [5] Nesargi S, Prakash R. *DADHCP: Distributed Dynamic Configuration of Hosts in a Mobile Ad Hoc Network*. University of Texas at Dallas, Department of Computer Science; Dallas, TX, USA: Jan, 2001. Technical Report UTDCS-04-01;
- [6] Patchipulusu P. *Dynamic Address Allocation Protocols for Mobile Ad Hoc Networks*. Texas A&M University; Dallas, TX, USA: Aug, 2001. Master's Thesis,
- [7] Mohsin M, Prakash R. IP Address Assignment in a Mobile *Ad Hoc* Network. Proceedings of Military Communications Conference (MILCOM); Anaheim, CA, USA. September 2002; pp. 856–861. Available online: <http://www.utdallas.edu/~ravip/papers/milcom02.pdf> (accessed on 25 November 2010).
- [8] Thoppian MR, Prakash R. A Distributed Protocol for Dynamic Address Assignment in Mobile *Ad Hoc* Networks. IEEE Trans. Mob. Comput. 2006;5:4–19.

Performance Metrics for Proactive and Reactive Routing Protocols in Mobile Adhoc Network

Basu Dev Shivhare¹, Anil Kumar Sajnani², Shalini Shivhare³

¹IT Deptt, Dehradun Institute of Technology, Greater Noida (U.P.)

²Assistant Professor, CS Deptt, AITTM, Amity University, Sector 125, Noida (UP)

³Research Engineer, Robosapiens Technologies Pvt. Ltd., Noida, (U.P)

¹basuiimt@gmail.com, ²asajnani@amity.edu, ³shalini@robosapiensindia.com

Abstract: A mobile ad hoc network (MANET) is a collection of mobile nodes that is connected through a wireless medium forming rapidly changing topologies. MANETs are infrastructure less and can be set up anytime, anywhere. The routing algorithms considered are classified into two categories proactive (table driven) and reactive (on demand). DSDV, DSR, and AODV algorithms are considered. We simulate various MANET routing algorithms in network simulator NS-2 and compare the performance metrics for each Routing protocol like throughput, packet delivery ratio and average end to end delay by varying the number of nodes (Node 20 and Node 30 Group) with different pause time by using cbr source traffic and node movement model.

Keywords: Mobile ad hoc network, pause time, throughput, packet delivery ratio, average end to end delay, NS2.

1. INTRODUCTION

Mobile ad hoc networks (MANET), that contain wireless mobile nodes, can freely and dynamically self organize into arbitrary and temporary ad hoc network topologies. Mobile MANET is a collection of communication devices or nodes that wish to communicate with infrastructure less support and without predetermined organization of available links. In MANET, Routing is main problem to route the data packets from one source node to destination node in networks.

The vision of mobile ad hoc networking is to support robust and efficient operation in mobile wireless networks by incorporating routing functionality into mobile nodes. Such networks are envisioned to have dynamic, sometimes rapidly-changing, random, multi hop topologies which are likely composed of relatively bandwidth-constrained wireless links.

2. SIMULATION STRATEGY

DSDV, DSR, AODV routing protocols can be implemented using Network Simulator (NS) 2.33. NS is a discrete event simulator targeted at networking research. NS 2 is an object oriented simulator, written in C++, with an OTcl interpreter

as a front-end. All the work is done under Linux platform, preferably Ubuntu.

A. Traffic and Movement

We can also define the traffic and movement pattern in separate files called CBR file and scenario file respectively. Cbr file can be created by using a tcl program called cbrgen.tcl which is present in the directory "ns-2/indep-utils/cmu-scen-gen/". To define the movement we use an exe file called setdest present in the folder "ns-2/indep-utils/cmu-scen-gen/setdest/".

B. Cbr file

This file is run with certain arguments to create the traffic connection file. The arguments are:

```
ns cbrgen.tcl [-type cbr/tcp] [-nn nodes] [-seed seed] [-mc connections] [-rate rate]
```

C. Scenario file

As cbr file is used to store the traffic connections, similarly scenario file is used to store the initial position of the nodes and movement of nodes at different pause times. Since it will be difficult to manually give initial position, movement of the nodes and their speed for each movement at different times we use a random file generator here also. The node movement generator is available under /indep-utils/cmu-scen-gen/setdest/ directory. It is available under the name "setdest", which is an exe file. This file is run with certain arguments to create the scenario file. The arguments are:

```
./setdest -n <num_of_nodes> -p pausetime -s <maxspeed> -t <simtime> -x <maxx> -y <maxy>
```

D. Simulation and Design

On the basis of results of *.nam file and *.tr file, the analysis is being done. We also evaluate the performance of these three routing protocols by taking number of nodes as a parameter with different pause time variations. NAM is a

built-in program in ns2-all-in-one package. It helps us to see the flow of packets between various nodes. With this, we are also able to know whether the packets have reached to their destination properly or dropped in between. NAM is invoked within the Tcl file. The NAM scripts are stored in *.nam file and scripts for tracegraph are stored in *.tr file.

The simulation is divided in SIX parts on the basis of number of nodes that vary:

1. DSDV with 20 nodes.
2. DSDV with 30 nodes.
3. DSR with 20 nodes.
4. DSR with 30 nodes.
5. AODV with 20 nodes.
6. AODV with 30 nodes.

With different pause time such as $p = 8.0$, $p = 16.0$, $p = 24.0$, $p = 32.0$, $p = 40.0$ with speed $M = 20.0$ m/s (constant)

The comparison of performance of DSDV, DSR, AODV based on the number of nodes is done on following parameters like Throughput, Packet delivery Ratio and average end-to end delay.

E. Simulation Environment

Parameter	Value
Simulator	ns-2 (Network Simulator 2.33)
Number of Nodes	20/30
Studied Protocol	DSDV, DSR, AODV
Simulation Time	3000 seconds
Simulation Area	1500 * 1500 m
Traffic Type	CBR(UDP)
Pause time	$P=8, 16, 24, 32, 40$ s
Speed	$M=20.0$ m/s (constant)
Channel Type	Wireless Channel
MAC Type	802.11
Traffic	Cbr
Packet Size	512 byte
Antenna Model	Omni

F. Performance Metrics:

The following different performance metrics are evaluated to understand the behavior of DSDV, DSR and AODV routing protocols

I-Throughput

II-Packet delivery ratio

III-The average end to end delay.

Throughput -Throughput is the total number of packets received by the destination.

Packet Delivery Ratio- The packet delivery ratio is defined as the number of received data packets (CBR) divided by the number of generated data packets (CBR).

Average end to end delay- The end to end delay is defined as the time a data packet is received by the destination minus the time the data packet is generated by the source.

The average time from the beginning of a packet transmission at a source node until packet delivery to a destination includes delays caused by buffering of data packets during route discovery, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer times.

3. SIMULATION RESULTS

A. Results of DSDV with pause time variation

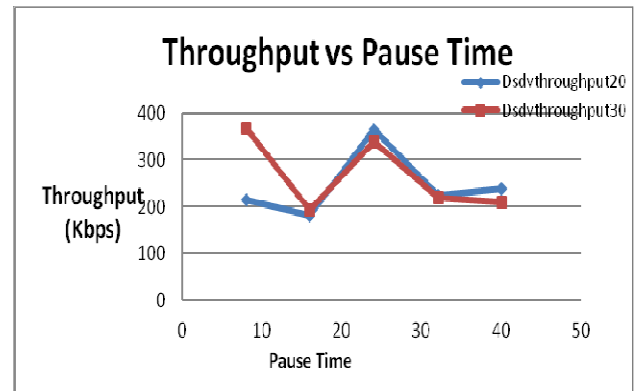


Fig. 1 (a) DSDV Throughput vs Pause Time variation

Initially for pause time $p = 8$ and $p = 16$, Throughput of DSDV routing protocol for node 20 is low than throughput of node 30. when pause time increases Throughput for node 20 increases than throughput of node 30.

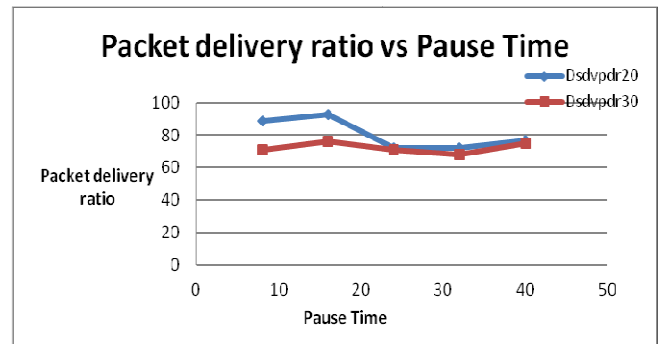


Fig. 1 (b) packet delivery ratio vs Pause Time variation

For pause time $p = 8, 16, 24, 32, 40$ sec. Packet delivery ratio of node 20 is superior than packet delivery ratio of node 30.

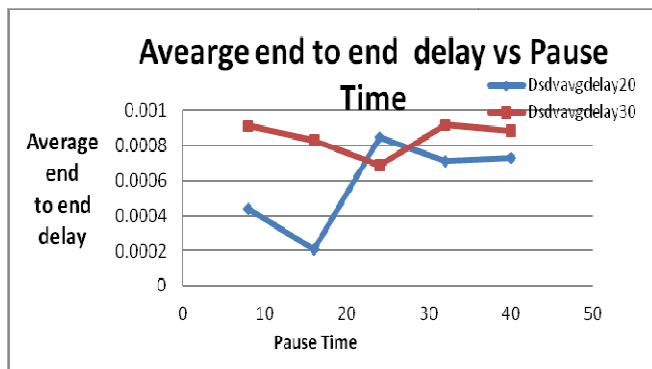


Fig. 1 (c) DSDV average end to end delay vs Pause Time variation

Average end to end Delay for packets increases as number of nodes increases.

B. Results of DSR with pause time variation-

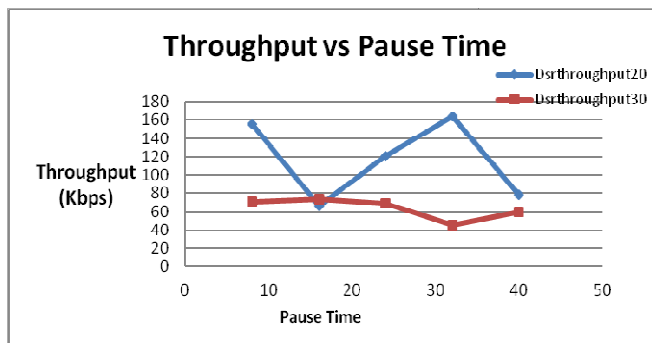


Fig. 2 (a) DSR Throughput vs Pause Time variation

When pause time increases, Throughput for fewer number of node is superior than throughput of bulky number of nodes.

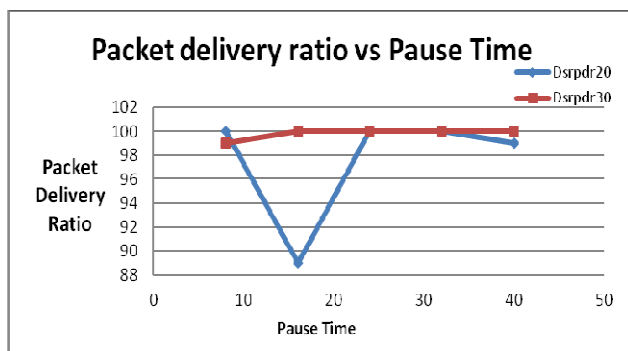


Fig. 2 (b) packet delivery ratio vs Pause Time variation

For pause time $p = 8$ sec Packet delivery ratio of node 20 is more than packet delivery ratio of node 30.

For pause time $p = 24, 32$ sec Packet delivery ratio of node 20 and node 30 is constant.

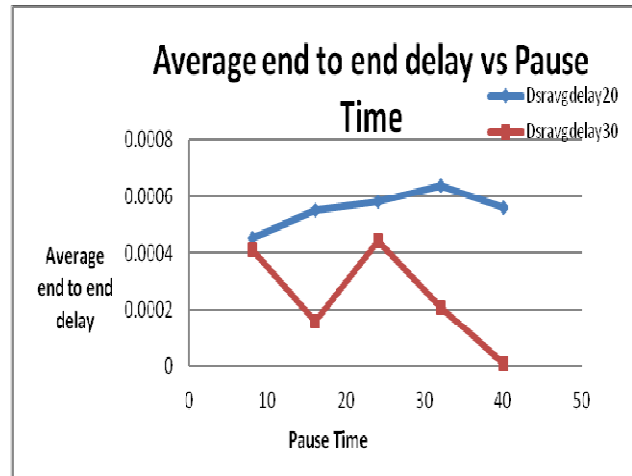


Fig. 2 (c) average end to end delay vs Pause Time variation

Average end to end Delay for DSR packets decreases as number of nodes increases. It means Average end to end Delay is low in DSR for bulky nodes.

C. Results of AODV with pause time variation-

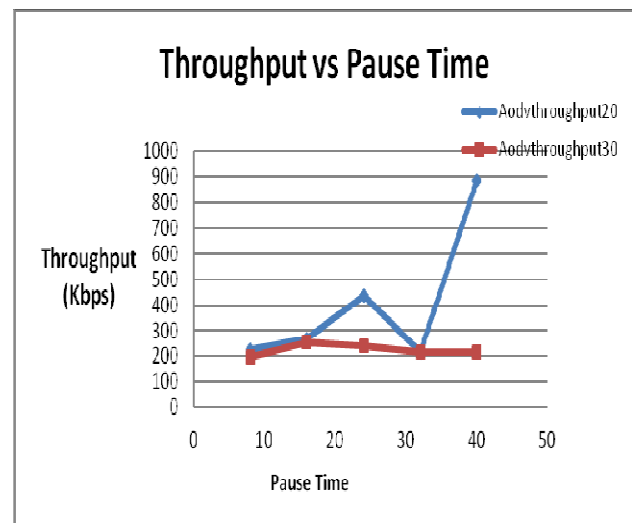


Fig. 3 (a) AODV Throughput vs Pause Time variation

For pause time variations Throughput of AODV routing protocol is better for fewer number of nodes than throughput of bulky number of nodes. When pause time increases Throughput for fewer number of node is high than throughput for bulky number of nodes.

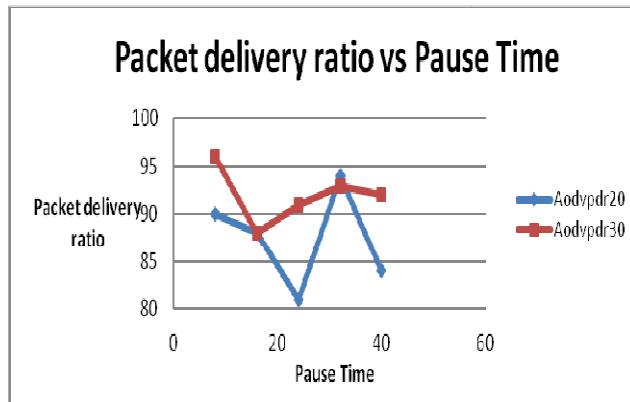


Fig. 3 (b) packet delivery ratio vs Pause Time variation

For pause time $p = 8, 16, 24, 32, 40$ sec Packet delivery ratio of node 20 is superior than packet delivery ratio of node 30.

When pause time increases Packet delivery ratio for bulky number of node is not superior than Packet delivery ratio of less number of nodes in AODV.

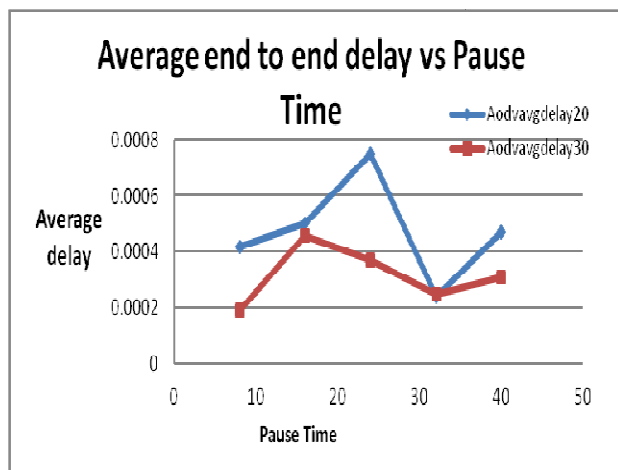


Fig. 3 (c) average end to end delay vs Pause Time variation

Average end to end Delay for AODV packets decreases as number of nodes increases. It means Average end to end Delay is low in AODV for large number of nodes for pause time $p = 8, 16, 24, 32, 40$ sec.

4. CONCLUSION

This paper does the realistic performance metrics of three routing protocols DSDV, AODV and DSR. The significant observation is, comparison results agree with expected results based on theoretical analysis.

Our analysis of the result guides us to conclude that:

A. DSDV Routing Protocol

Throughput-For Pause time variation, Throughput of DSDV Routing protocol for fewer number of nodes is superior than bulky number of nodes.

Packet Delivery Ratio- For Pause time variation, Packet delivery ratio of DSDV Routing protocol for small number of nodes is improved than large number of nodes.

Average End to End Delay- For Pause time variation, Average end to end delay of DSDV Routing protocol for fewer number of nodes is low than large number of nodes.

B. DSR Routing Protocol

Throughput- For Pause time variation, Throughput of DSR Routing protocol for less number of nodes is better than large number of nodes.

Packet Delivery Ratio- For Pause time variation, Packet delivery ratio of DSR Routing protocol for bulky number of nodes is enhanced than fewer number of nodes.

Average End to End Delay- For Pause time variation, Average end to end delay of DSR Routing protocol for less number of nodes is better than large number of nodes.

C. AODV Routing Protocol

Throughput- For Pause time variation, Throughput of AODV Routing protocol for more number of nodes is low than less number of nodes.

Packet Delivery Ratio- For Pause time variation, Packet delivery ratio of AODV Routing protocol for more number of nodes is better than less number of nodes.

Average End to End Delay- For Pause time variation, Average end to end delay of AODV Routing protocol for less number of nodes is better than more number of nodes.

REFERENCES

- [1] C. E. Perkins and P. Bhagwat "Highly Dynamic Destination Sequenced Distance-vector Routing (DSDV) for Mobile Computers, Proceedings of the ACM SIGCOMM '94 Conference.
- [2] Md. Anisur Rahman, Md. Shohidul Islam, Alex Talevski, "Performance Measurement of Various Routing Protocols in Ad-hoc Network", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009, IMECS 2009, March 18 - 20, 2009, Hong Kong.
- [3] Nor Surayati Mohamad Usop, Azizol Abdullah, "Performance Evaluation of AODV, DSDV & DSR Routing Protocol in Grid Environment", IJCSNS International Journal of Computer Science and Network Security, 2009.

- [4] S. A. Ade & P.A.Tijare, "Performance Comparison of AODV, DSDV, OLSR and DSR Routing Protocols in Mobile Ad Hoc Networks" International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 545-548
- [5] Sapna S. Kaushik & P. R. Deshmukh, "Comparison of effectiveness of AODV, DSDV AND DSR Routing Protocols In Mobile Ad Hoc Networks" International Journal of Information Technology and Knowledge Management, July-December 2009, Volume 2, No. 2, pp. 499-502
- [6] "Using Routing protocol property, comparison of proactive and reactive routing protocol in manet", IJETAE vol. 2, issue 3, march 2012, issn 2250-2459
- [7] NS2 Tutorial by Marc Greis

Wireless Sensor Networks with application to the Measurement and Detection of Air Pollution

Apurv Gupta¹, Rohit Kathait², Ankush Kapoor³

¹B.Tech 2nd Year ECE Department

Jawaharlal Nehru Government Engineering College, Sundernagar Distt. Mandi H.P

¹apurav.gupta7@gmail.com, ²rk.kathaitrohit@gmail.com

³Assistant Professor ECE Department

Jawaharlal Nehru Government Engineering College, Sundernagar Distt. Mandi H.P

ankush8818@yahoo.com

Abstract: Wireless sensors Networks are the need to present day world and they have attracted the attention of many researchers. The availability of low-cost hardware is enabling the development of wireless sensor networks (WSNs), i.e., networks of resource-constrained wireless devices that can retrieve multimedia content such as video and audio streams, measure temperature & pressure data from the environment. Wireless Sensors are becoming very popular in industrial processes for measurement and control, condition monitoring. They provide an easy, cost-effective path to redundancy without compromising safety. In this paper, the architecture of WSN is provided as well as its application to the measurement and detection of Air Pollution is being discussed.

Keywords: Wireless Sensor Networks (WSNs), Low power design, Healthcare monitoring.

1. INTRODUCTION

Wireless sensor network (WSNs) is basically a collection of sensing devices that can sense, process and talk to their peers. Need for a thorough discussion on Wireless Sensor Networks (WSNs) had been felt considering the rapid progress in the research, development and deployment of WSNs. The wireless networking capability of the sensor enabled nodes, have resulted in various interesting applications ranging from surveillance, smart homes, precision agriculture, disaster detection, underwater, to vehicular and supply chain management applications. Driven by technology advances in low-power networked systems and medical sensors, we have witnessed in recent years the emergence of wireless sensor networks (WSNs) in healthcare[1]. Wireless Sensor Networks has been emerging from the vision of Smart dust project in 1998 that required enabling both communication and sensing capabilities in order of cubic millimeter. The *Sensor Node*, which is a basic element of Wireless Sensor Network, is composed of Sensing, Computation and wireless Communication unit. These sensor nodes are hence capable of observing physical phenomenon, process the observed and received information and communicate the observed or processed information to the nearby sensor nodes to form a network of sensor nodes

called Wireless Sensor Networks (WSNs) [2]. These WSNs carry the promise of drastically improving and expanding the quality of care across a wide variety of settings and for different segments of the population. For example, early system prototypes have demonstrated the potential of WSNs to enable early detection of clinical deterioration through real-time patient monitoring in hospitals, enhance first responders' capability to provide emergency care in large disasters through automatic electronic triage, improve the life quality of the elderly through smart environments, and enable large-scale field studies of human behaviour and chronic diseases.

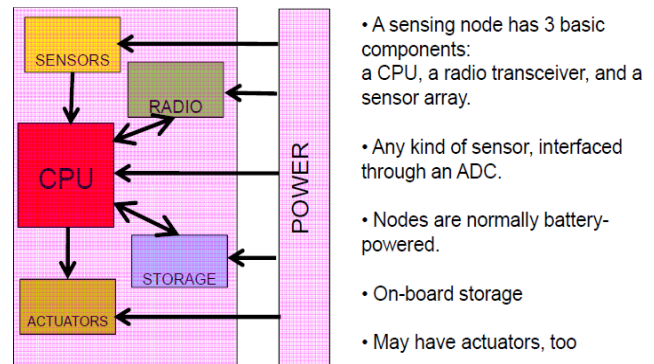


Fig. 1. Wireless Sensor Network

The main characteristics of a WSN include:

1. Power consumption constrains for nodes using batteries or energy harvesting.
2. Ability to cope with node failures
3. Mobility of nodes
4. Dynamic network topology
5. Communication failures

6. Heterogeneity of nodes
7. Scalability to large scale of deployment

2. FEATURES OF WSN'S IN OUR APPLICATION REGIME

The features enabling a broad range of applications for air pollution sensing networks are described as:

A. Ocean sampling network

Networks of sensors and AUVs, such as the Odyssey-class AUVs, can perform synoptic, cooperative adaptive sampling of the atmosphere [2]. Experiments such as the Monterey Bay field experiment demonstrated the advantages of bringing together sophisticated new robotic vehicles with advanced air models to improve the ability to observe and predict the characteristics of the atmosphere.

B. Environmental monitoring

Air sensor networks can perform pollution monitoring (chemical, biological and nuclear). For example, it may be possible to detect the chemical slurry of antibiotics, estrogen-type hormones and insecticides to monitor the atmosphere [3]. Monitoring of winds, improved weather forecast, detecting climate change, understanding and predicting the effect of human activities on the atmosphere, are other possible applications.

C. Disaster prevention

Sensor networks that measure atmospheric activity from remote locations can provide tsunami warnings to us.

D. Assisted navigation

Sensors can be used to identify hazards on the seabed, locate dangerous rocks or shoals in mountainous region and to perform bathymetry profiling.

E. Mine reconnaissance

The simultaneous operation of multiple AUVs with acoustic and optical sensors can be used to perform rapid environmental assessment and detect mine-like objects.

3. ARCHITECTURE OF WSN

Upto now plenty of researches in WSN have been conducted, and formed several main research platforms, the architecture of WSN commonly used by most of the research platforms as shown in Fig1. The sensor nodes randomly dispersed in the monitoring area by aircrafts spreading, manually deployed rockets ejecting, constitute a network through self organization method. Each of these nodes has

the capability of collecting data and routes data back to the base stations and the base stations send the information to the center through internet. The end users can browse and process data through internet from the center. A data acquiring unit, a processing unit, a data transceiver unit and a power unit. Data acquiring unit is usually composed of two sub units: a sensor and an analog to digital converter (ADC) [3]. The selection of sensors lies on the interesting objects. The analog signals of observed phenomenon produced by the sensors are converted into digital signals by the ADC, and then sent to processing unit. The processing unit, which is generally associated with a small storage sub-unit or sometimes with an application sub unit, manages the procedure that makes the sensor node collaborate with the other nodes to carry out the assigned sensing task. The transceiver unit connects the sensor node to the wireless sensor network. One of the

Important components of the sensor node are the power unit. The power unit may include a power management sub-unit that helps the processor monitor and manage the energy consumption [4]. As the lifetime of a WSN is so important, a management of power that can effectively protract the lifetime available is obligatory. Most WSN routing techniques and sensing tasks require the knowledge of location with high accuracy. An operating system (OS) is needed for the sensor node's managing the power, scheduling the services, sensing a task, tracking a target, and communicating with the neighboring nodes.

4. CHARACTERISTICS OF WSN:

The main characteristics of a WSN include:

- Power consumption constrains for nodes using batteries or energy harvesting.
- Ability to cope with node failures
- Mobility of nodes
- Dynamic network topology
- Communication failures
- Heterogeneity of nodes
- Scalability to large scale of deployment
- Ability to withstand harsh environmental conditions
- Ease of use
- Unattended operation
- Power consumption

Sensor nodes can be imagined as small computers, extremely basic in terms of their interfaces and their components. They usually consist of a processing unit with

limited computational power and limited memory, sensors or MEMS (including specific conditioning circuitry), a communication device (usually radio transceivers or alternatively optical), and a battery. Other possible inclusions are energy harvesting modules, secondary ASIC and possibly secondary communication devices. The base stations are one or more components of the WSN with much more computational, energy and communication resources.

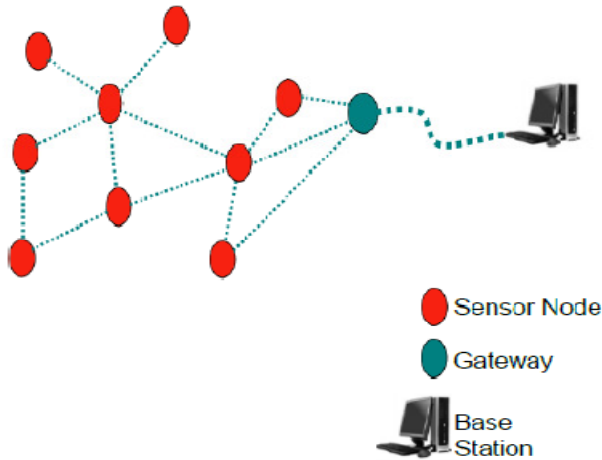


Fig. 1. Architecture of Wireless Sensor Networks

They act as a gateway between sensor nodes and the end user as they typically forward data from the WSN on to a server. Other special components in routing based networks are routers, designed to compute, calculate and distribute the routing Sensor Node Gateway Base Station tables. Many techniques are used to connect to the outside world including mobile phone networks, satellite phones, radio modems, long-range Wi-Fi links etc.

5. ROUTING IN WSNs:

Recent advances in WSNs have it possible for small, inexpensive, low-power sensor nodes to be distributed across a geographical location. The information can be gathered and sent to the end user or BS through wireless communication. These tiny sensor nodes have sufficient intelligence for signal processing and data broadcasting, compared to other wireless networks, WSNs have resource constraints such as limited battery power, bandwidth and memory. WSNs have to periodically self-organize and generate routes from the nodes to the BS. The sensor nodes broadcast information to communicate with each other in the network.

The performance metrics usually considered when working with a WSN are power consumption, connectivity, scalability and limited resources. Sensor nodes play the dual role of data collection and routing and therefore need energy

to perform. Malfunctioning of few nodes, due to hardware or lack of energy, in the network will cause significant topology changes, leading to greater consumption of energy and rerouting of packets. Therefore, energy efficient schemes and communication protocols are being designed for WSNs.

6. APPLICATIONS OF WSNs

WSN may consist of many different types of sensors such as seismic, low sampling rate magnetic, thermal, visual, infrared, acoustic and radar, which are able to monitor a wide variety of ambient conditions that include temperature, humidity, vehicular movement, lightning condition, pressure, soil makeup, noise levels, the presence or absence of certain kinds of objects, mechanical stress levels on attached objects, and the current characteristics such as speed, direction, and size of an object. WSN can be used for continuous sensing, event detection, and local control of actuators. Applications of WSN can be categorized into many fields such as military, environmental, healthy, home, commercial, and industrial areas.

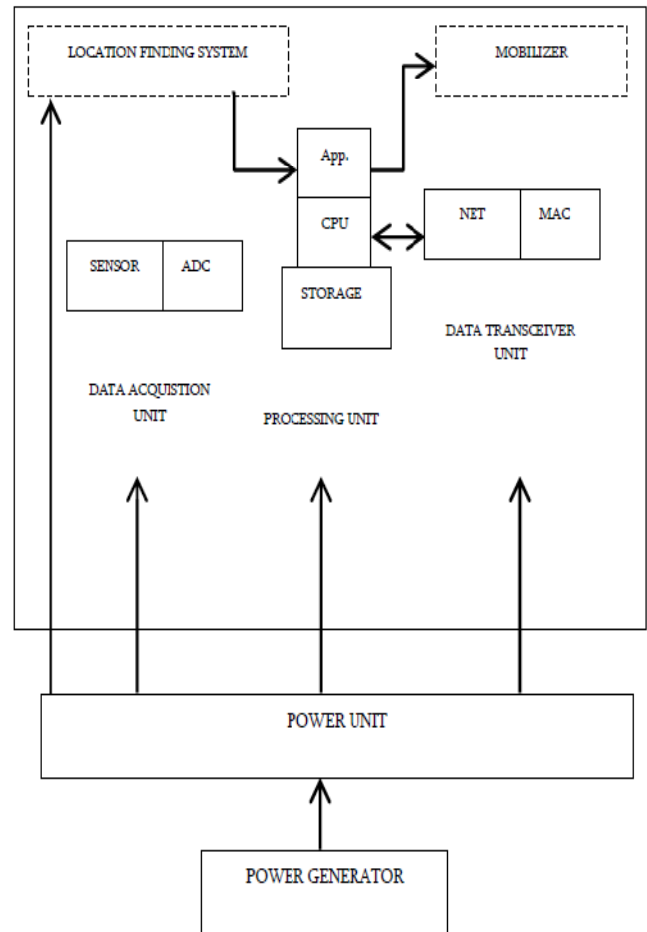


Fig. 2. Components of a node in Wireless Sensor Networks

A. MILITARY APPLICATIONS:

WSN can be an integral part of military Command Control, Communications, Computing, Intelligence, Surveillance, Reconnaissance and Targeting systems and is competent for monitoring friendly forces, equipment and ammunition, and other tasks. The rapid deployment, self-organization and fault diagnosis characteristics of WSN make it a very promising sensing technique for these military applications.

B. INDUSTRIAL APPLICATIONS:

Industrial applications include robot control and guidance in automatic manufacturing environments; industrial process control and automation; smart structure with sensor nodes embedded inside; machine diagnosis; factory instrumentation; local control of actuators; instrumentation of semiconductor processing chambers; monitoring or rotating machine; monitoring of wind tunnels and anechoic chambers and etc.

C. ENVIRONMENTAL APPLICATIONS:

Environmental applications include tracking the movements of birds, small animals and insects; monitoring environmental conditions that affect crops and livestock irrigation; forest fire detection flood detection etc. In forest fire detection, since millions of sensor nodes can be strategically, randomly, and densely deployed in a forest, integrated using RF Optical systems, sensor nodes can relay the exact origin of the fire to the end users before the fire is spread uncontrollable.

7. WAPMS: THE PROPOSED AIR POLLUTION MONITORING SYSTEM

A Wireless Sensor Network Air Pollution Monitoring system proposes the wireless air pollution monitoring systems (WAPMS). This network comprises of a number of sensor nodes and a data network which passes the information to the base stations. The system can be used to send the commands to the nodes and also allows the nodes to gather data autonomously.

The strategies to deploy the WSN were as follows:

1. The region was divided into number of several smaller areas this improved management and coordination.
2. Single cluster head deployed in each region.
3. Randomly deploy sensor nodes in the selected regions.
4. A number of sinks used to collect the data.

Air quality index was used in WAPMS which is an indicator of air quality based upon the amount of air pollutants in the air.

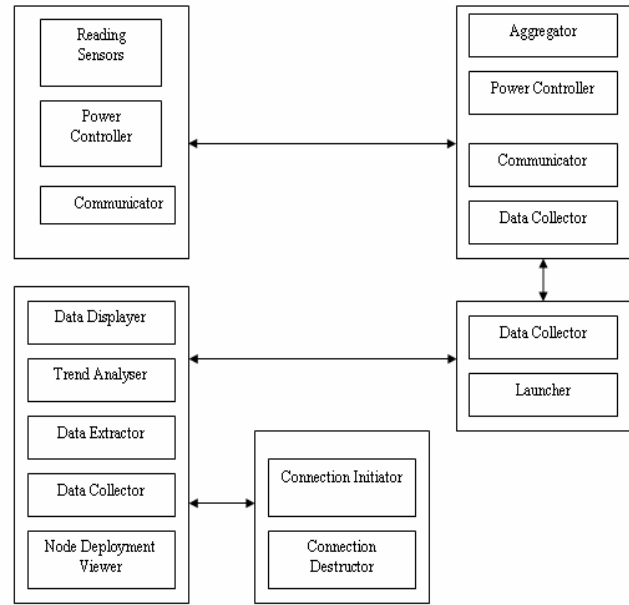


Fig. 3. Architecture diagram of the system

A. Simulation and Results

WAMPS was simulated using the JiST/swans stimulator it runs over a standard virtual java machine. SWANS is a scalable wireless network stimulator built atop the JiST platform making it an independent software. When given an area and date, the system displays the corresponding AQI readings and the health concern in that area as show in the fig below. WAMP also allows timely monitoring of the data.

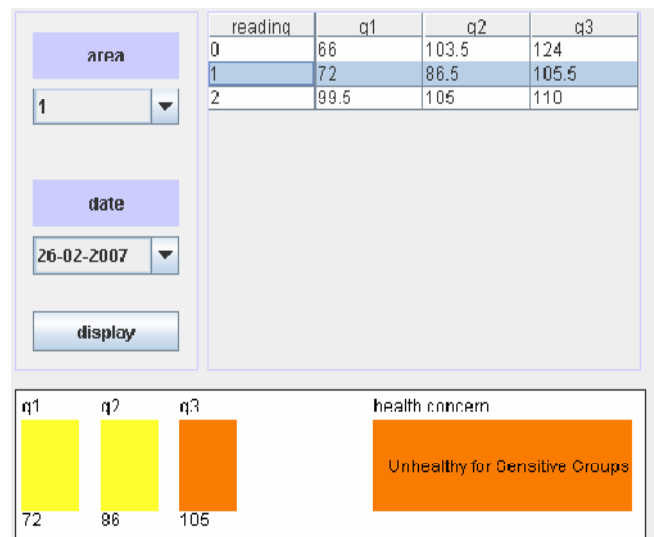


Fig. 4. AQI for the selected area

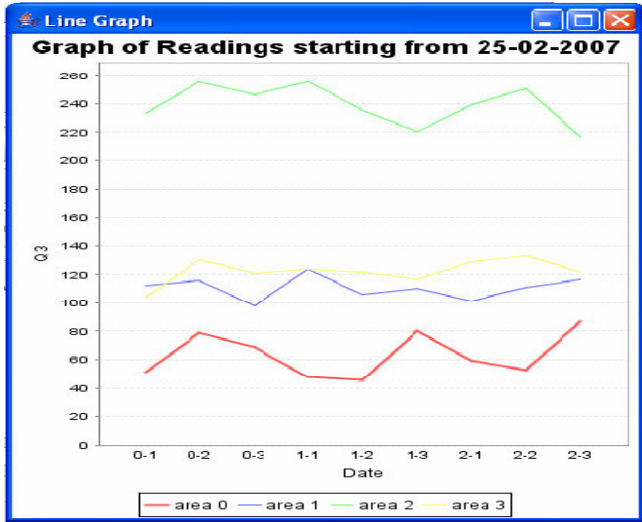


Fig. 5. Line graph for that area

8. DYNAMIC GAS SENSOR NETWORK FOR AIR POLLUTION MONITORING:

In paper [4] the use of dynamic gas sensors proposed for air pollution monitoring. A dynamic network was proposed and calibration was done by comparing sensor outputs. If the gas sensors were placed on running vehicles on the street they will move randomly at some frequency two or more sensors share same location. The calibration is done by comparing the sensor output at that moment.

A. Simulation

The area is divided into 10 by 10 mesh grids as shown in the figures below. The gas sensors move in only four directions (up, down, right, and left) and are fixed initially. They provide accurate gas concentration values all the time.

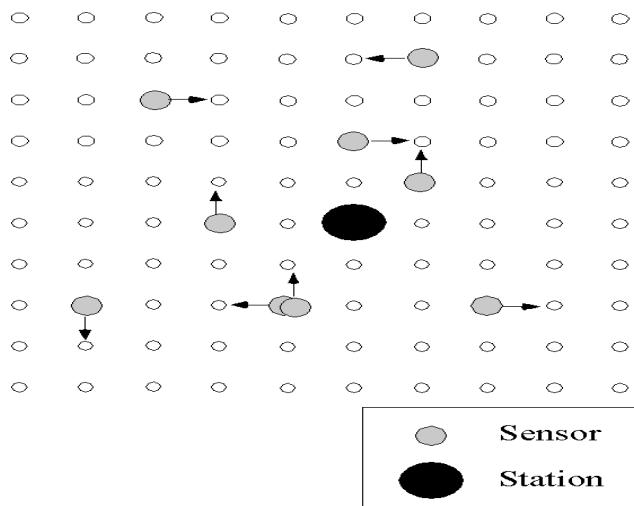


Fig. 6: Simulation Setup

9. CONCLUSIONS

The major motivation behind our study and the development of the system is to help the government to devise an indexing or monitoring system to categories air pollution.

The results shown in this work correspond to the working test of the network and sensor nodes. Advanced processing based on multiple-input-single-output neural networks is implemented at the network sensing nodes to obtain temperature and humidity compensated gas concentration values.

Further development in sensor networks could represent a shifting data flow, storage, and communication. Technological improvements in processing capacity and form factor will enable nodes to decide on-the-fly whether to process, compress, store, act on, or send sensor data back to users. As such, implementation of WSNs for equipment condition monitoring does not require the knowledge of an RF engineer or designer.

REFERENCES

- [1] Ruizhong Lin, Zhi Wang and Youxian Sun, "Wireless Sensor Networks Solution for Real Time Monitoring of Nuclear Power Plant" in Proceedings of the 5th World congress in Intelligent control and Automation, June 15-19, 2004.
- [2] JeongGil Ko, Chenyang Lu, Mani B. Srivastava, John A. Stankovic, "Wireless Sensor Networks for Healthcare".
- [3] Grefory Chen, Scott Hanson, David Blaauw, Dennis Sylvester, "Circuit Design Advances for Wireless Sensing Applications"
- [4] Feng Zhao, Leonidas Guibas, "Wireless Sensor Networks – An information processing approach".
- [5] Jagannathan Sarangapani, "Wireless Ad Hoc and Sensor Networks – Protocols, Performance and Control"

Routing of Autonomous Wheeled Mobile Robot in External Environment using Wireless Sensor Network

Anuj Chadha

*Institute of Diploma Studies, Nirma University, Ahmedabad, India
anuj.chadha@nirmauni.ac.in*

Abstract: This paper proposes the application of wireless sensor network for routing of wheeled mobile robots in dynamic external environment. Wireless sensor network (WSN) in navigation of mobile robots avoids the need of knowledge of map in advance. Sensor nodes will enable mobile robot to build the map of the external location while moving around. The problem of localization can be solved by determining robot's physical distance from sensor nodes. In proposed navigation system, the robot can navigate autonomously without the need for a map, by acquiring the information from RF emission sensors deployed in an external environment. This approach is easy to implement and proposed model shows that accurate navigation can be achieved by this method.

Index Terms: Wireless Sensor Network (WSN); QoS; Sensor Nodes; Automated Guided Vehicle (AGV)

1. INTRODUCTION

Advancements in microelectronics field have led to significant reduction in size and cost of components required for sensor nodes. This has led to rapid development in the field of wireless communication. Large number of sensor nodes connected with neighboring nodes and subsequently to control room forms a Wireless Sensor Network. Sensor nodes are capable of computation as well as communication with neighbor nodes. Wireless sensor network has to be power efficient. Efficiency can be increased by using many application specific routing algorithms.

Wireless Sensor network has found its application in many areas such as in military area for surveillance and battlefield assessments, structure monitoring, pollution and toxic level monitoring, rainfall and flood monitoring, wind speed along with its direction and temperature monitoring to forecast weather conditions, habitat monitoring, medical applications like tracking patients, drug administration monitoring in the hospitals, traffic monitoring and vehicle identification. Despite the innumerable applications of WSNs, these networks have several restrictions, e.g. limited energy supply, limited computing power, and limited bandwidth of the wireless links connecting sensor nodes. One of the main design goals of WSNs is to carry out data communication

while trying to prolong the lifetime of the network and prevent connectivity degradation by employing aggressive energy management techniques [1]. The design of routing protocols in WSNs is influenced by many challenging factors like energy consumption, node deployment, fault tolerance, coverage, data aggregation, Quality of Service (QoS) [2].

Indoor routing of mobile robot is comparatively easier task than outdoor routing. Navigation of robots in external environment is a far complex process because robots are not aware about their current position. The problem of localization and path planning can be solved by continuous communication with neighboring sensor nodes. Sensor network in external environment has many constraints like battery life, coverage and quality of service. Hence, efficient routing algorithms have to be implemented to increase the life time of sensor nodes.

The problem of mobile robot navigation in external environment deals with navigation on the scale of a few meters around the desired outdoor location, such as desert, wide open area, where the main task for successful implementation lies in the robot's current location, which means to determine an assignment of coordinates to nodes in either a wireless sensor network [3]. There are other method of position measurement of the robots e.g. odometry, Inertial Navigation, Magnetic Compasses, Global Positioning System (GPS) [4]. The position measurement method using sensor network is consistent with accurately measured pair wise node distance. The localization problem in mobile robots has received attention of many researchers in recent years. This problem can be solved by building occupancy grid map of the present environment around the robot [4, 11]. This solution can be used to determine the navigation direction such that the robot is safely guided towards a desired location. The user in the control room should also acquire position of each sensor node if sensor nodes are not stationary. If correct position of sensor nodes is not known, information obtained from nodes is useless. The sensor node must know its own position first, and then it can know the position where the events happen, in an effort to fulfill the

tasks such as location and tracking for the object. Combination of more than one orthogonal technique for path planning helps to reduce possible errors in position estimation. The mobile robots will be able to communicate with each other and with the wireless sensor network. Whenever multiple mobile robots are operating in nearby region, their motions have to be coordinated to avoid deadlocks or collisions. The coordination between multiple robots will be handled by central controller [5]. The location information has been proven to be useful in the remote surveillance, router communication, object tracking and network administration [10].

In my propose method for routing of mobile robots in outdoor environment, robot can find the routes autonomously without the need of the map. Of course, prior knowledge of map helps robot to navigate but map cannot be reliable for navigation in dynamic external environment. Hence, robot can navigate autonomously by extracting information from radio wave sensor installed in outdoor environment and the relative inter node distance between them. Once enough information about inter node distances to draw a graph of the sensor node is obtained, we can run any of several well suited localization schemes to compute node coordinates, then by measuring the distance from one sensor node after another using triangular localization method, the robot locates itself and knows its current orientation. Compared with the traditional methods, my proposed algorithm is conceptually simple and easy to implement.

The other topics of the paper are as follows: Part 2 deals with brief review of related work of the localization methods for mobile robots. Part 3 describes the system design of our proposed model in external environment with provided Sensor Network. Part 4 describes operation of proposed model and real time applications of our model in outdoor environment. Part 5 consists of conclusions and future work.

2. RELATED WORK IN LOCALIZATION METHOD

The first step of localization is knowledge of the exact position of the vehicle. Automated guided vehicles (AGVs) should not be confused with mobile robots because AGVs use magnetic tape, buried guide wires or painted strips on the ground for routing. AGVs cannot freely alter their path. On the other hand mobile robots can easily alter their path according to sensor outputs. Hence, positioning of mobile robots has no elegant solution. The best solution is to use GPS based mobile robot positioning for outdoor environment [9]. The reduction in size, cost and power dissipation level of various sensors and microcomputers due to advancements in microelectronics field has made sensor network feasible to be used in external environment. All localization problems can be seen as a problem attempting to find the position of unknown nodes using as much of the available information as possible.

Recently, various relative methods have been proposed for solving this larger problem. The RADAR system [7] based on RF signal strength measurements can track the location of users within an indoor environment. The Cricket location support system [8] is proposed for indoor localization by using ultrasound information. In indoor routing of mobile robots [5], a rectangular grid of beacons and RF connectivity constraints are considered for indoor localization. The detailed analysis on performance of the system can be seen in [10]. An iterative multi lateration is investigated in [11]. The presented algorithm performs well when a large percentage of beacons are available, the graph connectivity is high and precise range measurements can be determined.

As mentioned above, most approaches generally rely on the assumption that each node being able to know its distance to the others close to it. According to type of measurement for the angle or distance of the sensor node information, the Localization algorithms can be roughly divided into two categories: range-based and range-free [5].

In range-based algorithms, sensor nodes estimate their distance to seeds using some specialized hardware. These measurements are used in methods like triangulation [4] or trilateration [4], which are based on the concept that a node location can be uniquely specified when at least the coordinates of three reference points are available for a node. Although the use of range measurements results in a fine grained localization scheme, range-based algorithms require the sensors contained hardware to make range measurements. From the sensor node point of view, range-based algorithms are computationally expensive, which limits their use as a tool for practical use. Later attempts have been made to overcome this problem, the recently proposed range-free localization method provides a solution to the drawbacks of range-based method, which has been proven to be suitable for the sensor network.

Range-free algorithms do not use radio signal strengths, angle of arrival of signals or distance measurements and do not need any special hardware. Range-free algorithms require that each node knows the following items:

- i. Which nodes are within radio range
- ii. Their location estimates
- iii. The (ideal) radio range of sensors.

No other information is required for robot localization.

Thus, range-free techniques are more computational efficient for the reason that they do not require sensors to be equipped with any special hardware, but use less information than range-based algorithms.

Considering the pros and cons of the traditional approaches, and our outdoor environment analysis, our proposed localization method can be described as follows:

- i. The measurement for the sensor nodes orientation acquisition and detection
- ii. The estimation of the position based on the acquired orientation information.
- iii. The final localization based on the revolutions.

The system design, operation of environment analysis and detailed description of the algorithm will be given in the following section.

3. SYSTEM DIAGRAM

The proposed system is the design of a wireless sensor network with both static and mobile nodes to monitor environmental parameters. A set of self navigating robots acts as mobile nodes in the system. Each robot is embedded with obstacle finding sensors and environmental parameter monitoring sensors. There is a centralized server to coordinate the activities of robots. Total area considered for surveillance is divided into a number of regions. Central controller will assign operating region of each robot and hence it has to balance the sensing load and provide surveillance of the whole area.

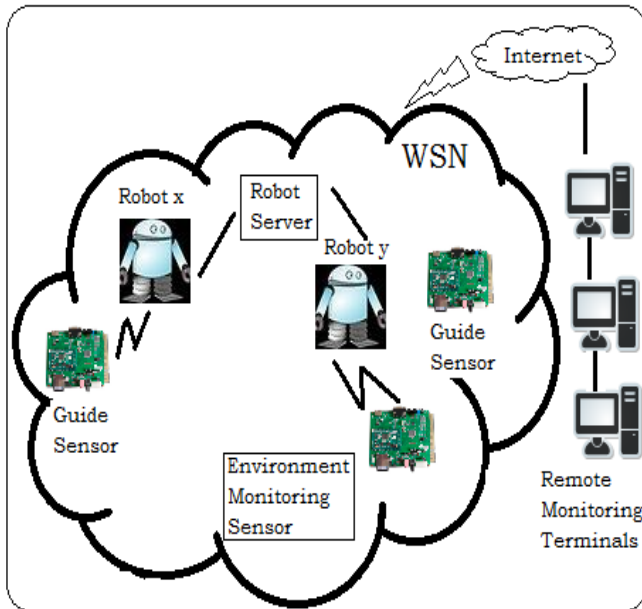


Fig. 1. System Design

The data captured by mobile robots is monitored at remote monitoring terminals by connecting them with robot servers in the Wireless sensor network via internet.

4. PROPOSED MODEL

A. Environment Navigation Analysis

In proposed navigation system for mobile robot, sensor nodes will act as signpost and accordingly they inform the robot of the next node to pass through, while the robot follows each sensor node along the routing path and interchange distance information with each sensor node.

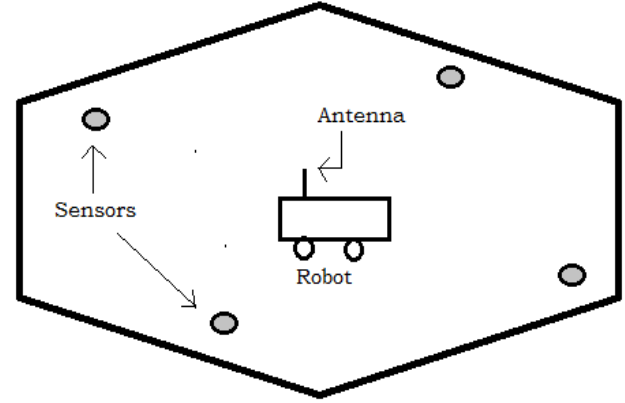


Fig. 2. Positioning with triangulation in outdoor environment

Fig. 2 indicates the valid coverage of the sensors. As we can see from the figure, the sensor nodes are scattered around the robot's operation environment and form the whole ad-hoc sensor network, their positions are known. The radio emission devices are set on each sensor node, which can emit the electro-magnetic wave with specific frequency. The robot is equipped with a rotatable antenna, which is used for receiving the radio signals from the sensor nodes, and labeling each sensor by extracting the frequency information. We denote the wireless sensor nodes set by $S = \{S_i\}$, and sensor nodes groups $N_i = \{(S_{i1}, S_{i2}, S_{i3})\}$, and we denote the corresponding frequency groups by $F_i = \{(F_{i1}, F_{i2}, F_{i3})\}$, as well as frequency space $F = \{F_i\}$. We divide the mobile robot's operation space E into a series of subspaces E_i where E_i belongs to $E = \{E_i\}$. We set one sensor nodes group in every subspace E_j , where each sensor node will be fixed at the different position. Note that the position $P_i = (P_1, P_2, P_3)$ is known in advance, which is easy to implement.

When mobile robot attempts to navigate in the operation space E , it will receive the radio signals from the various wireless sensor nodes scattered cross the environment. The antenna will rotate by a specific speed, during the location procedure, and the antenna will search for the specific frequency in the frequency space F , after a certain number of rotation cycles, when the magnitude M_i of frequency for the specific orientation exceeds the predefined threshold T_i , which means one sensor node has been detected with the known orientation. The robot traverses the environment by the limited number of search, and all the sensors in the

subspace sensor nodes group N_i and the relative orientations can be detected.

We build the map between the sensor nodes and its positions as follows,

$$\text{Map} = m(N_i, E_i, P_i, F_i) \quad (1)$$

In above equation, m stands for the corresponding relations among the sensor nodes, subspaces, frequency bands and robot's absolute position.

B. Positioning with triangulation

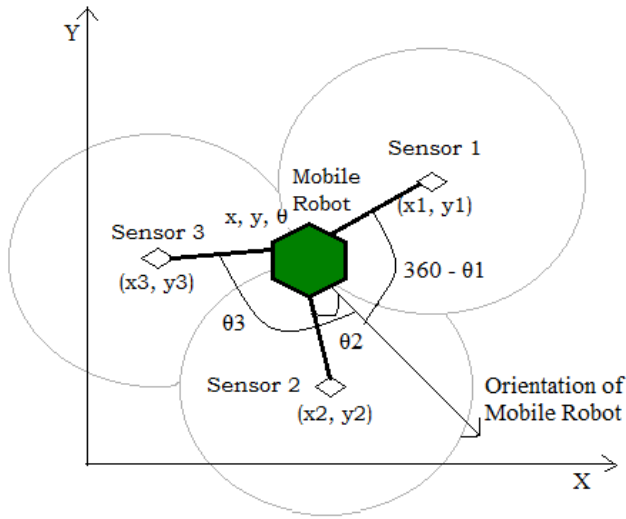


Fig. 3. Triangulation

In this paper, triangulation method is employed to estimate position of mobile robot using information of radio sensors in the sensor group N_i in subspace E_j , including their a priori coordinates in the reference frame and directions with respect to robot. As shown in Fig. 3, mobile robot detects three sensors S_i ($i=1, 2, 3$) around it.

C. Real Time Applications

Mobile robots following proposed model for path planning and navigation can be used to replace the battery / sensor node in the wireless sensor network. This model can be used in any external environment where position determination is impossible with conventional process [12].

D. Conclusion and Future Work

This paper proposes a new model of mobile robot positioning based on wireless sensor network. With the help of radio sensors deployed with known coordinates in the outdoor environment, mobile robot can detect the directions

of sensors and estimate its position by triangulation method. The future work will focus on the improvement on robustness and extend my work to more general environment. My focus will also be on to use lightweight algorithms for sensor nodes to communicate with each other efficiently. Due to which, life time of the sensor node will be increased.

REFERENCES

- [1] Heinzelman W., Chandrakasan A. and Balakrishnan H. (2000), "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proceedings of the 33rd Hawaii International Conference on System Science, pp 8020.
- [2] Bagga P, Chadha A. (2010), "Design Issues in routing algorithms for Wireless Sensor Network", Proceedings of the National Conference on Emerging Vistas in 21st century, pp 55-58.
- [3] B Jamal N. Al-Karaki Ahmed E. Kamal "Routing Techniques in Wireless Sensor Networks: A Survey" Dept. of Electrical and Computer Engineering Iowa State University, Ames, Iowa, Volume 53 Issue 7, pp 945-960.
- [4] Borenstein J., Everett H.R., Feng L., Wehe D., "Mobile Robot Positioning – sensors and techniques", Invited paper for the Journal of Robotic Systems, Special Issue on Mobile Robots. Vol. 14 No. 4, pp. 231 – 249.
- [5] Siyao Fu, Zeng Guang Hou and Guosheng Yang "An indoor navigation system for Autonomous mobile robot using Wireless Sensor Network", Proceedings of IEEE International Conference on Networking, Sensing and Control – 2009.
- [6] Joshua D Freeman, Simi S, "Remote Monitoring of Indoor environment using mobile robot based wireless sensor network", 6th International Conference on Computer Science and Education – 2011.
- [7] P. Bahl, V. Padmanabhan, RADAR: An in-building RF-based user location and tracking system, in Proc. of Info com, 2: 775-584, 2000.
- [8] N. Priyantha, A. Chakraborty, H. Balakrishnan, The cricket location-support system, in Proc. of International Conference on Mobile Computing and Networking, (Boston, MA), pp. 32-43, 2000.
- [9] N. Bulusu, J. Heidemann, D. Estrin, GPS – low cost outdoor localization for very small devices, IEEE Personal Communications Magazine, vol. 7, pp. 28-34, Oct. 2000.
- [10] N. Bulusu, J. Heidemann, D. Estrin, and T. Tran, Self-configuring localization systems: Design and experimental evaluation, ACM Transactions on Embedded Computing Systems (ACM TECS), Special issue on networked embedded systems, 2003.
- [11] C. Savarese, J. M. Rabaey, J. Beutel, Localizing in distributed ad-hoc wireless sensor networks, in Proc. Of ICASSP/OI, 4: 2037-2040, 2001.
- [12] Gakuhari H., Songmin J., Hada Y., Takase K., "Real Time Navigation for Multiple Mobile Robots in a dynamic environment", Proceedings of the 2004 IEEE conference on Robotics, Automation and Mechatronics, Singapore, December, 2001.

A Discussion on Hardware Platforms for Wireless Sensor Network

Malang Shah¹, Saurabh Mehta²

*Electronics & Telecommunication Engineering Department
Vidyalankar Institute of Technology, Mumbai, India
malangshah, saurabh.mehta}@vit.edu.in*

Abstract—Nodes also known as motes are the basic building blocks of a Wireless Sensor Network (WSN). There are various motes available today, ranging from application specific to the generic type. WSN network consists of simple cost effective nodes which are densely deployed in the field and are controlled by several gateway nodes. This paper compares weC, Rene, Dot, Mica, Telos, Sunspot, Tmote Sky and Wasp mote with respect to the hardware components used, processing capabilities, RF communication characteristics, power consumption etc. The WSN history, its applications, and design challenges are discussed briefly. The need for operating system (OS) for WSN is identified and a comparison of Tiny OS, Contiki, MANTIS, Nano RK and LiteOS is presented. A discussion on few of the many open issues in hardware platforms like multi radio architecture and multiple directional antennas is presented which is a new design concept in the evolution of hardware platforms for WSN.

Index Terms—Wireless Sensor Nodes, Wireless Sensor Network (WSN), History of WSN, Applications of WSN, Design Challenges, RF transceivers and Operating Systems.

1. INTRODUCTION

A Wireless Sensor Network (WSN) is a system consists of small; battery powered computing devices called Wireless Sensor Nodes or Motes. A mote measures physical, chemical, electrical, and optical parameters for e.g. temperature, humidity, pressure, light, etc. Then it transmits these parameters to other motes over a radio or optical link. Generally a WSN is set up by deploying few hundreds to thousands of motes across a region to be monitored. A region is then divided into several clusters each having few hundreds of low cost nodes. Each cluster has a base station node or gateway node to monitor simple nodes. WSN is used in range of applications like Area monitoring, Structural Health Monitoring, Agriculture, Military, Home Automation, Target tracking, etc [1].

The paper is divided into 7 sections. Section 1 is introduction which explains what is meant by a WSN. Section 2 presents a brief history of WSN. Section 3 illustrates and explains the blocks within a generic mote and features of various commercial motes are explored along with a comparison chart. Section 4 explains Wireless Sensor

Network and its applications. Section 5 highlights the main design challenges of a WSN and in section 6, a comparison of the embedded operating systems like Tiny OS, Contiki, MANTIS, Nano RK and LiteOS is presented. The final section 7 is a discussion on the open issues in hardware platforms for WSN and future design trends.

2. HISTORY OF WSN

Modern research on sensor networks started around 1980 with the Distributed Sensor Networks (DSN) program at the Defense Advanced Research Projects Agency (DARPA) [2]. Technology components for a DSN were identified in a Distributed Sensor Nets workshop in 1978. These included sensors (acoustic), communication, processing techniques and algorithms (including self-location algorithms for sensors), and distributed software (dynamically modifiable distributed systems and language design). WSN is an outcome of DARPA funded academic research project during 1990s [3]. Mainstream research was started by the University of California, Berkeley, USA. The goal was to create a tiny autonomous computer called sensor nodes, which observes their environment with built in sensors and report back these events to a remote base station. In their early applications, hundreds or thousands of these nodes were deployed for monitoring environmental conditions by periodically sending the data to a remote base station.

UC Berkeley developed the early commercial mote WeC in the year 1998[4], based on Atmel's 8 bit microcontroller AT90LS8535, operating at 4 MHz with 512 Bytes of RAM, 8K Bytes of Flash and 32K Bytes of EEPROM. RFM TR1000 Radio transceiver was used, operating at 915 MHz and was supporting a Bandwidth of 10 Kbps. With the development of low power microcontrollers and Radio chips, the battery operated devices has seen a tremendous demand in the past decade. The low power consumption of these semiconductor chips allows a device to be operated on a battery providing an operational life that can range from few months to several years. This has led to the development of many commercially available motes that became popular in the last decade between the years 1998 to 2007. Few of the popular commercial motes are WeC, Rene, Dot, Mica,

Telos and Tmote Sky. These motes have been widely used in both academic research and industrial applications.

3. COMMERCIAL MOTES

Generally a mote consists of few sensors (transducers), a microcontroller, an RF transceiver, an actuator if required and a power source (battery) as illustrated in Fig.1. A *Sensor* is basically a transducer which converts the physical, chemical, mechanical, biological or magnetic parameters into electrical signals [5]. These electrical signals are analog in nature and converted into digital format by an A/D converter of a *Microcontroller* so as to process this information and store it in an inbuilt memory or on a remote computer. An important function of a microcontroller is to communicate with RF Transceiver, which transmits and receives radio signals. It is used for wireless communication between motes. The electrical signals from microcontroller are converted into radio signals and vice versa. The power supply to a mote is either from a non rechargeable or rechargeable type *battery* depending on the application. Other sources for power are mains supply, solar energy, etc. An *actuator* is used for performing a controlling action for e.g. opening or closing a valve. An electromechanical relay is an example for actuator.

Commercial motes are classified as special purpose, generic and gateway type of nodes [12]. A gateway node is different than the simple and generic type of nodes in terms of processing capabilities which are enhanced by higher data bus width, memory capacity and throughput of the microcontroller. A gateway collects data sensed by individual motes of a cluster and performs complex functions like data compression and forwarding data to legacy networks like internet, PSTN, Cellular, LAN etc. Table I is a comparison chart of few of the commercial motes like WeC, Rene, Dot, Mica, Telos, Sunspot, Tmote Sky and Wasp mote. Table II groups the different motes, having common hardware or firmware features considering microcontroller, memory, radio transceiver and OS used [4].

4. WIRELESS SENSOR NETWORK AND APPLICATIONS

A WSN consists of many motes. An application may require several hundreds of motes to be deployed across a given area. There are two types of WSNs: structured and unstructured [5]. An unstructured WSN is densely populated with motes, which are placed in ad hoc manner into the field. While in a structured WSN all or several motes are well positioned according to a pre plan. A base station or a sink node collects data from all the motes within its range and acts as a gateway to connect the WSN to another network infrastructure for e.g. Internet as shown in Fig.2. Thus a WSN can be monitored from a remote location using a computer. A gateway node may not be within a range of

every mote in a network, in such a case the sending mote uses other intermediate motes along the path that leads to the gateway node. A data can be transferred between a mote and a gateway using multiple paths following different nodes.

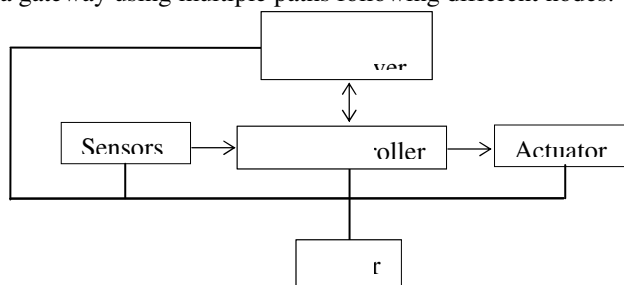
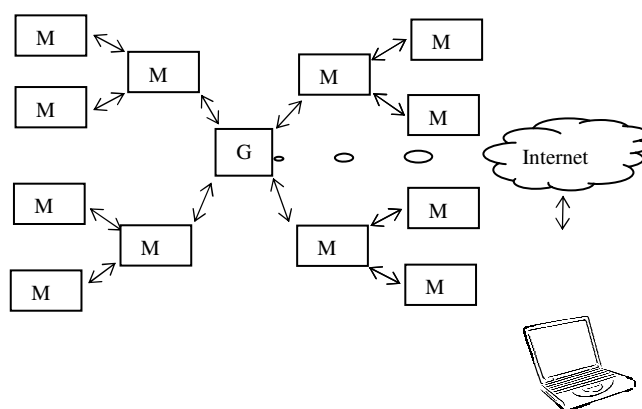


Fig. 1. Basic blocks within a Mote



Legends used, M- mote, G - Gateway or Base station
Fig. 2. A typical Wireless Sensor Network

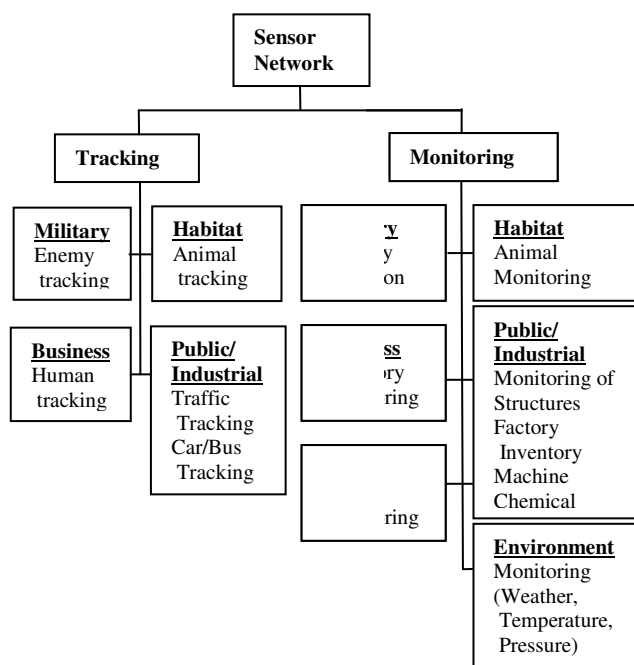





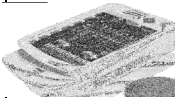




Fig. 3. Overview of Sensor Applications [5]

Wireless Sensor Networks has been widely used in Tracking and Monitoring applications. An overview of sensor applications is illustrated in Fig. 3 [5]. WSN is used for Structural health monitoring [6] of bridges like Alamosa Canyon Bridge, NM, USA, Di Wang Tower, Guangdong, China and Geumdang Bridge, Icheon, Korea. For Habitat monitoring a network was deployed on Great Duck Island, a small island off the coast of Maine for monitoring Leach's Storm Petrel [7]. For Environment monitoring a wireless

sensor array was deployed in July 2004 at Volcan Tingurahua, an active volcano in central Ecuador to monitor volcanic eruptions with low-frequency acoustic sensors [8]. Intel's Wireless Vineyard is used for agricultural monitoring. In this application, the network is expected not only to collect and interpret data, but also to use such data to make decisions aimed at detecting the presence of parasites and enabling the use of the appropriate insecticide [9].

Table I: Comparison of Commercial nodes

Sr #	Parameters	weC [6],[11]	Rene [6],[11],[13]	Dot [6],[11], [20]	Mica [6], [10]	Telos [11]	Sunspot [13], [15]	Tmote Sky [14], [15], [16]	Waspnode [16]
1	Image								
2	Computation Technology								
2.1	Micro-controller	AT90LS853	Atmega163L	Atmega163L	ATmega103L	TI-MSP 430 family	AT91RM920T	TI-MSP 430 family	ATmega1281
2.2	Bus (Bits)	8				16	32	16	8
2.3	Clock Speed (MHz)	4			4	8		8	
2.4	Prog. Memory (KB)	8	16		128	48	4000	48	128
2.5	Data Memory (KB)	32		32	512	1024	512	10	2 GB SD Card
2.6	ADC Resol.(bits)	10					12		10
3	Wireless Communication Technology								
3.1	Transceiver	TR1000				CC 2420		CC 2420	
3.2	Frequency (MHz)	868 / 916				2400			868/ 900/ 2400
3.3	Modulation		on-off key			O-QPSK		DSSS-QPSK	
3.4	Data Rate (Kbps)	10			40	250		250	
4	Operating System								
4.1	OS				Tiny OS				
5	Power Source and Consumption								
5.1	Battery	Coin Cell, 575 mAh	3V/2850 mAh		3V/2850 mAh				3.3 V - 4.2V
5.2	Active (mW)		24		100	41			
5.3	Idle (μ W)				60				
6	Miscellaneous Parameters								
6.1	Year	1999	2000	2001	2002	2004		2006	
6.2	Manufacturer	UC Berkeley- Crossbow		UC Berkeley	UC Berkeley- Crossbow	UC Berkeley	Sun Microsystems	Moteiv	
6.3	Cost in USD		100				750		
6.4	On Board Sensor							5	Temperature & Accelerometer
6.5	ID chip				DS2401	48-bit chip			Present
6.6	Range (m)				60			Outdoor:125 Indoor: 50	500 / 700/ 10,000
6.7	Dimensions	2.5 x 2.5 x 1.3 cm						2621 mm ²	73.5 x 51 x 13 mm

5. DESIGN CHALLENGES FOR WSN

Design issues in WSN are broadly classified and illustrated in Fig. 4[5]. Sensor technology refers to the type of sensors or transducers required for an application. Since WSN consists of several hundreds to thousands of nodes, and each node will have many types of transducers, hence it is essential to use small size, low cost sensor technology. With the advancements in Micro Electro-Mechanical System (MEMS) technology it is now possible to produce low cost, small size, smart sensors. Each sensor node is an individual system. In order to support different application software on a sensor system, operating systems, and storage schemes are needed.

The Communication Protocols enables communication between the application and sensors and also between the sensor nodes. Communicational protocols are required at each of the five layers of the TCP/IP reference model, which are physical, data - link, network, transport and application layer. After deploying nodes, they should self configure themselves to create a network and thereby can be scaled to any size. To achieve this functionality network management protocols called services are needed. Services are developed to enhance the application and to improve system performance and network efficiency. Localization, Coverage, Security, Synchronization, Data aggregation and Cross -layer Optimization are issues under Services category.

Table II: List of motes based on common design components

Sr#	Particulars	Motes	Features	Limitations	Remark
1	Microcontroller				
1.1	Atmel AT90LS8535	WeC, Rene 1	Clock - 4 MHz RAM / Flash - 512 / 8K	8 bit	
1.2	Atmel Atmega 163	Rene 2, Dot	Clock-8 MHz RAM / Flash - 1K / 16K		
1.3	Atmel Atmega 128L	Mica, BT Node, Mica2, Mica2Dot, iBadge, CENS Medusa MK2, Nymph, MicaZ, AquisGrain, DSYS25, Ember RF Module, Module, Fleck	Clock - 8 MHz RAM / Flash - 4K / 128K		
1.4	Atmel AT91FR40162S	Sun Spot	Clock - 75 MHz Flash - 256K		
1.5	TI MSP430F1611	TelosB / Tmote Sky, eyesIFXv2, SHIMMER	Clock - 8 MHz Flash / EEPROM - 10K		
2	External Memory				
2.1	EEPROM 32K	WeC, Rene 1, Rene 2, Dot			
2.2	EEPROM 48K	TelosB / Tmote Sky, eyesIFXv2, SHIMMER			
2.3	EEPROM 512K	Mica, BT Node, Mica2, Mica2Dot, iBadge, CENS Medusa MK2, Nymph, MicaZ, AquisGrain, DSYS25, Ember RF Module, Module, Fleck			
2.4	EEPROM 2M	Sun Spot			
3	Radio Transceiver				
3.1	RFM TR1000	WeC, Rene 1, Rene 2, Dot, Mica	BW (Kbps) - 10 Freq (MHz) - 916.5	Low data Rate	
3.2	Chipcon CC 1000	BT Node, Mica2, Mica2Dot, Nymph	BW (Kbps) - 38.4 Freq (MHz) - 900		
3.3	Chipcon CC 2420	Telos, MicaZ, BSN Node, AquisGrain, Pluto, iMote2, XYZ Sensor Node, ProSpeKz II	BW (Kbps) - 250 Freq (MHz) - 2400		
4	Operating System				
4.1	TinyOS	WeC, Rene 1, Rene 2, Dot, Mica, BT Node, Spot ON, Telos, Mica2, Mica2Dot, iMote, Spec, MicaZ, CIT Sensor Node, BSN Node, AquisGrain, TelosB/Tmote Sky			
4.2	Squawk VM (Java)	Sun Spot			
5	Multiple Radios	BT Nodes			Future Work
6	Multiple Directional Antenna & MAC protocol	None			

As sensor nodes operate on limited battery power, energy usage is a very important design consideration in a WSN; and there has been significant research focused on harvesting and minimizing energy. Solar power is also used for recharging the battery of nodes to increase life time of the network. Other options for energy sources are electromagnetic energy harvesting, fuel cells, etc. Coverage of an area with WSN depends on application requirement. For an application with high degree of accuracy of the sensed data, more number of nodes is required for dense deployment. In such scenarios less power is required to communicate with immediate nodes due to short distance between them. But for an application that requires nodes to be placed at larger distances, more power is required for radio communication. Thus protocols are required that manages power usage optimally and increase life time.

Cross layered approach is more energy efficient as compared to the traditional individual layered approach of the protocol stack design. In the cross layered approach, the protocol stack is treated as a system and not individual layers, independent of each other. Layers share information from the system. Thus the overhead is reduced. Security in

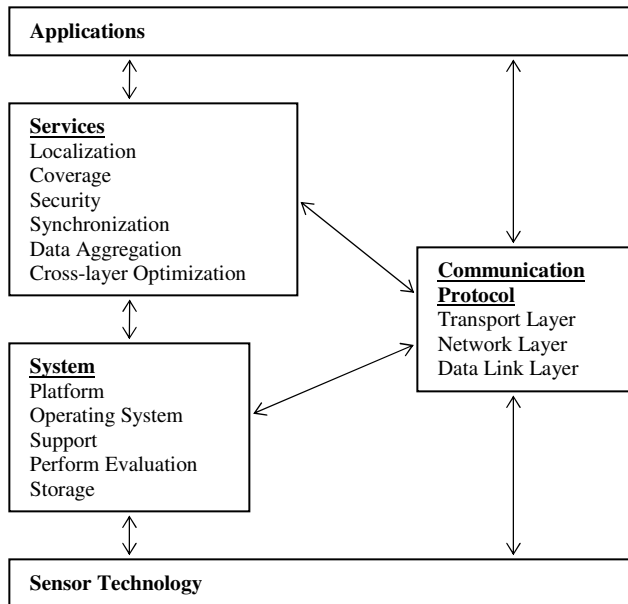


Fig. 4. Broad classification of various issues in a WSN [5]

WSN is also a concern, since nodes are operating on license free bands hence unauthorized access, eavesdropping are possible; protocols are required for protection against such threats. There are many challenges at each stages of designing a wireless sensor network. As this paper is focused on review of hardware platforms hence the following section discusses design issues faced while using embedded operating systems, dedicated to WSN. An

operating system is an integral part of a hardware platform and is required for implementing the functionalities of WSN on small memory footprint devices due to the limitations posed by scarce memory and processing capabilities of low cost microcontrollers.

6. EMBEDDED OPERATING SYSTEMS FOR WSN

A sensor node hardware is simple but used in range of applications from simple data gathering to understanding complex physical phenomenon like estimation of location of a source in a region, which is monitored using gas sensors for detection of harmful gases e.g. Carbon Monoxide, Hydrogen Sulphide, etc. An operating system (OS) bridges the gap between simple hardware of the node and complex requirements of the application and it plays a central role in building scalable distributed applications that are efficient and reliable. The basic functionalities of an OS include resource abstractions for various hardware devices; interrupt management and task scheduling, concurrency control, and networking support [18]. There are many popular OS available for WSN, few of them are Tiny OS, Contiki, MANTIS, Nano-RK and Lite OS. Table III is a comparative summary of these operating systems [19]. While designing OS for wireless sensor network following factors are important design considerations.

• Architecture

The architecture of an OS has an influence on the size of the OS kernel as well as on the way it provides services to the application programs. Some of the well known OS architectures are the monolithic architecture, the micro-kernel architecture, the virtual machine architecture and the layered architecture.

• Programming Model

There are two popular programming models provided by typical WSN OSs, namely: event driven programming and multithreaded programming. Multithreading is the application development model most familiar to programmer, but in its true sense rather resource intensive, therefore not considered well suited for resource constraint devices such as sensor nodes. Event driven programming is considered more useful for computing devices equipped with scarce resource but not considered convenient for traditional application developers.

• Scheduling

The Central Processing Unit (CPU) scheduling determines the order in which tasks are executed on a CPU. In traditional computer systems, the goal of a scheduler is to minimize latency, to maximize throughput and resource utilization, and to ensure fairness. The selection of an

appropriate scheduling algorithm for WSNs typically depends on the nature of the application. For applications having real-time requirements, real-time scheduling

algorithm must be used. For other applications, non-real-time scheduling algorithms are sufficient.

Table III: Operating Systems Summary

OS/ Feature	Architecture	Programming model	Scheduling	Memory Management and Protection	Communication Protocol Support	Resource Sharing	Support for Real- time Applications
TinyOS	Monolithic	Primarily event Driven, support for TOS threads has been added	FIFO	Static Memory Management with memory protection	Active Message	Virtualization and Completion Events	No
Contiki	Modular	Protothreads and events	Events are fired as they occur. Interrupts execute w.r.t. priority	Dynamic memory management and linking. No process address space protection.	<i>uIP and Rime</i>	Serialized Access	No
MANTIS	Layered	Threads	Five priority classes and further priorities in each priority class.	Dynamic memory management supported but use is discouraged, no memory protection.	At Kernel Level COMM layer. Networking Layer is at user level. Application is free to use custom routing protocols.	Through Semaphores.	To some extent at process scheduling level (Implementation of priority scheduling within different processes types)
Nano-RK	Monolithic	Threads	Rate Monotonic and rate harmonized scheduling	Static Memory Management and No memory protection	Socket like abstraction for networking	Serialized access through mutexes and semaphores. Priority Ceiling Algorithm is implemented	Yes
LiteOS	Modular	Threads and Events	Priority based Round Robin Scheduling	Dynamic memory management & it provides memory protection to processes.	File based communication	Through synchronization primitives	No

- *Memory Management and Protection*

In a traditional operating system, memory management refers to the strategy used to allocate and de-allocate memory for different processes and threads. Two commonly used memory management techniques are static memory management and dynamic memory management. Static memory management is simple and it is a useful technique when dealing with scarce memory resources. At the same time, it results in inflexible systems because run-time memory allocation cannot occur. On the other hand, dynamic memory management yields a more flexible system because memory can be allocated and de-allocated at run-time.

- *Communication Protocol Support*

In the OS context, communication refers to inter-process communication within the system as well as with other nodes in the network. WSNs operate in a distributed environment, where sensor nodes communicate with other nodes in the network. All WSN OSs provide an Application Programming Interface (API) that enables application program to communicate. It is possible that a WSN is composed of heterogeneous sensor nodes; therefore the communication protocol provided by the OS must also consider heterogeneity. In network-based communication, the OS should provide transport, network, and MAC layer protocol implementations.

• Resource Sharing

The responsibility of an OS includes resources allocation and resource sharing, which is of immense importance when multiple programs are concurrently executing. The majority of WSNs OSs today provide some sort of multithreading, requiring a resource sharing mechanism. This can be performed in time e.g., scheduling of a process/thread on the CPU and in space e.g., writing data to system memory. In some cases, we need serialized access to resources and this is done through the use of synchronization primitives.

7. DISCUSSION & FUTURE WORK

WSN is a new research area that has potential applications in structural health monitoring, body area networking, pervasive computing, data collection from remote areas like forest, mountains, glaciers, wildlife monitoring, habitat monitoring, early detection of calamities, etc. These are few examples of WSN but can be extended to making any things smarter, intelligent and networked with the internet for easy access and monitoring. Hardware platforms for WSN has several open issues that needs to be addressed for making the technology suitable for the wide range of applications. These open issues are briefly discussed in this section.

As WSN requires nodes in bulk, the size and cost of the nodes need to be improved[15]. Silicon manufacturers need to integrate more functions on same chip thereby reducing the cost further. They will play an important role in the evolution of WSN technology. Hardware should operate at very low voltages for reducing the size of the components and by lowering power consumption, which will inturn reduce the size of the battery and lower the cost of the system. Along with the hardware, the MAC components, software design also need to improve for increasing energy efficiency.

Energy consumption can be reduced with low power devices as well as by using software methods i.e. using energy efficient algorithms. One of the techniques for reducing power consumption is to use multiple directional antennas instead of single omnidirectional antennas at each node. As radiation from an omnidirectional antenna is more than the directional type of antenna, using multiple directional antennas will reduce power consumption significantly and thus increases lifetime of the network. This technique will also reduce interference from network nodes but it increases the complexity in the MAC design. Another area open for research is the use of cognitive radio technology in WSN. A node with multiple radios is also a new design concept. Such architecture has been implemented in BT Nodes [5], where Bluetooth technology was used and one radio was configured as a master while the other radio of the same node was acting as a slave. They were used to form a multi-hop network.

8. CONCLUSION

We have reviewed the history of WSN, along with general block diagram of motes. A typical architecture of WSN is illustrated using multi-hop technique. A comparison is made between various popular commercial motes. Few application examples are listed for the WSN which are deployed on the field. Design challenges for the WSN are discussed in brief followed by the design considerations for embedded operating systems. A comparative study of various OS is also presented. Two new techniques for designing motes using multiple directional antennas and multi radios are also discussed in brief.

REFERENCES

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, *Wireless sensor networks: a survey*, Computer Networks, Elsevier, 2002.
- [2] Chee-Yee Chong, Srikanta P. Kumar, *Sensor Networks: Evolution, Opportunities, and Challenges*, Proceedings of the IEEE, vol. 91, No. 8, August 2003.
- [3] Raja Bose, *Sensor Networks—Motes, Smart Spaces, and beyond*, Pervasive Computing, IEEE CS, July – September 2009.
- [4] Michael Healy, Thomas Newe and Elfed Lewis, *Wireless sensor node hardware: a review*, IEEE Sensors 2008 Conference.
- [5] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, *Wireless sensor network survey*, Elsevier Computer Networks 52 (2008), pp. 2292–2330.
- [6] Jerome P. Lynch and Kenneth J. Loh, *A summary review of wireless sensors and sensor networks for structural health monitoring*, The Shock and Vibration Digest, Vol. 38, No. 2, March 2006, pp. 91–128.
- [7] Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, John Anderson, *Wireless sensor networks for habitat monitoring*, WSNA'02, September 28, 2002, Atlanta, Georgia, USA.
- [8] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, "Monitoring volcanic eruptions with a wireless sensor network," in *Proceedings of the Second European Workshop on Wireless Sensor Networks (EWSN'05)*, Jan. 2005.
- [9] Puccinelli, D.; Haenggi, M., *Wireless sensor networks: applications and challenges of ubiquitous sensing*, Circuits and Systems Magazine, IEEE, Volume: 5, Issue: 3, Publication Year: 2005, pp. 19 – 31.
- [10] Jason L. Hill, David E. Culler, *MICA: a wireless platform for deeply embedded networks*, IEEE Micro, November December 2002.
- [11] Joseph Polastre, Robert Szewczyk, and David Culler, *Telos: enabling ultra-low power wireless research*, Fourth International Symposium on Information Processing in Sensor Networks, 2005.
- [12] Jason Hill, Mike Horton, Ralph Kling, and Lakshman Krishnamurthy, *The platforms enabling wireless sensor networks*, Communications of the ACM - Wireless sensor networks, Vol. 47, Issue 6, June 2004, pp. 41-46.

- [13] Vini Madan and SRN Reddy, Review of wireless sensor mote platforms, VSRD International Journal of Electrical, Electronics & Comm. Engg., vol.2, 2012.
- [14] Jan Beutel, Metrics for sensor network platforms, REALWSN '06 Uppsala, Sweden, ACM, 2006.
- [15] Ana-Bele'n Garcí'a-Hernando, José'-Ferna'n Martí'nez-Ortega, Juan-Manuel Lo'pez-Navarro, Aggeliki Prayati, Luis Redondo-Lo'pez, MsC, Problem solving for wireless sensor networks, Springer.
- [16] <http://www.libelium.com/products/waspmote>
- [17] Thang Vu Chien, Hung Nguyen Chan and Thanh Nguyen Huu, "A comparative study on hardware platforms for wireless sensor networks", International Journal on Advanced Science Engineering Information Technology, vol. 2, (2012) No. 1.
- [18] Wei Dong, Xue Liu, Providing OS support for wireless sensor, networks: challenges and approaches, IEEE communications surveys & tutorials, vol. 12, N.. 4, fourth quarter 2010.
- [19] Muhammad Omer Farooq and Thomas Kunz "Operating systems for wireless sensor networks: a survey", *Sensors* 2011.
- [20] <http://www.atmel.com/Images/doc1142.pdf>

An Overview of Localization Techniques and Algorithms for Wireless Sensor Network

Sudhir P. Kasar¹, Ashish K. Shekhar², Malang Shah³, Saurabh Mehta⁴

^{1,2,3,4}Dept. of Electronics and telecommunication

Vidyalankar Institute of Technology, Wadala, Mumbai-400037, India

¹sudhir.kasar, ²ashish.shekar, ³malangshah, ⁴saurabh.mehta}@vit.edu.in

Abstract: Wireless Sensor Network (WSN) is collection of sensor nodes which are autonomous devices with integrated sensing, processing, and wireless communication capabilities. WSNs are used in variety of applications such as environmental monitoring, rescue, transportation, military, agriculture, etc. For such applications, localization of nodes is an important design criterion. Mechanism of location discovery and establishing spatial relationship among sensor nodes is known as localization. Localization techniques are classified under numerous parameters. Based on data computing, there are two types of localization techniques Centralized and Distributed localization. This paper reviews various localization algorithms on the basis of positioning principles, techniques used, advantages, limitations, challenging issues and various applications.

Index Terms: Wireless Sensor Networks (WSN), Nodes, Localizations algorithm, Centralized and distributed localization techniques

1. INTRODUCTION

Wireless Sensor Networks (WSN) is a new paradigm of measuring and controlling physical phenomena. WSNs are self organizing, self manageable and co-operative networks which are battery operated i.e. limited power source. It is composed of tiny elements called nodes which communicate to the sink node or base station through multi-hop communication link. The placement of nodes can be in structured or unstructured manner. For both the cases the nodes need to find the neighboring nodes through which it can reach to the base station. These nodes also need to know their spatial position in network. The process of finding relative spatial position for effective communication is called *Localization*.

This paper is divided into V sections. Section I is an Introduction of WSN and also defines what is meant by localization. Section II explains the basic elements of localization. Section III describes generalized classification of the localization techniques. Section IV is a comparative study of Centralized and Distributed localization techniques [1] followed by the concluding section on the open issues in Localization.

2. BASIC ELEMENTS OF LOCALIZATION

Figure 1 depicts the elemental configuration of localization system in WSN [4]. There are three components of this system described briefly below.

1. *Distance estimation:* This component provides the angular and range information between two adjacent nodes through various techniques such as Received Signal Strength Indicator (RSSI); Angle of arrival (AOA), Time of Arrival (ToA), etc.

2. *Position Computation:* This component calculates the actual position of nodes with respect to some reference node triangulations, multilateral and trilateral are some of the basic techniques to compute the position.

3. *Localization algorithm:* This algorithm helps some or all the nodes to know their relative position with respect to all other nodes in network. This element performs the task on the data supplied by position computation element. Directed position estimation and ad-hoc positioning system are some of the famous algorithm used for localization.

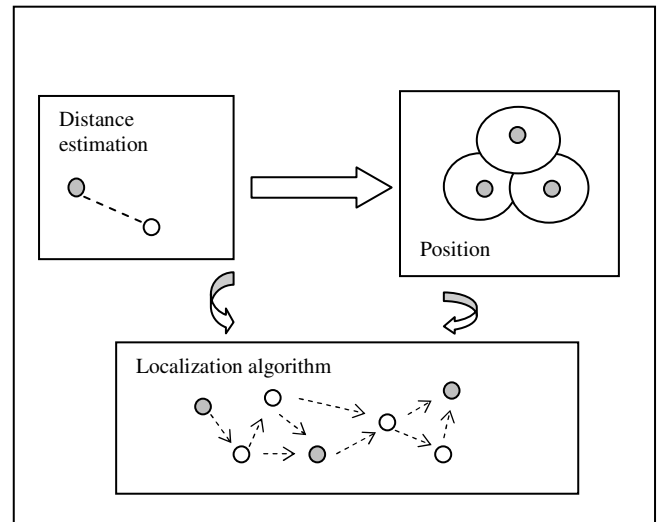


Fig. 1. Elements of Localization [4].

3. CLASSIFICATION OF LOCALIZATION TECHNIQUES

Localization techniques are classified under numerous parameters, few of the classifications are as follows:

1. Based on data computing: *Centralized and distributed*
2. Based on Reference node positioning: *Anchor free and Anchor based*
3. Based on coverage distances: *Range free and Range based*
4. Based on Mobility: *Mobile and Stationary*

The focus of study lies in the first class of localization algorithm i.e. based on localization data computing. In Centralized algorithm the position data is being computed at node and processed at the power full centralized base station whereas in distributed algorithm the position data is computed and processed in each node itself. In the centralized algorithm, the overload of computing position data at each node is reduced which improves the efficiency of network [1]. The disadvantage of this algorithm is the wastage of power due to the back and forth of position data which needs to be carried out several times. While in case of distributed algorithm there is no need of back and forth data flow towards a central entity. The disadvantage of this algorithm is in the wastage of lots of energy and time on computing position and neighbors' discovery if periodic location updates are used [1]. Classification of each algorithm is shown in Fig. 2 [1].

4. LOCALIZATION CATEGORIES

Centralized Localizations can be categorized as follows.

1. **MDS-MAP (Multi-dimensional scaling)** is an algorithm that uses the law of cosines and linear algebra to reconstruct the relative positions of the points based on the pair wise distances [1]. It consists of three steps. In the first step the scheme computes shortest paths between all pairs of nodes in the region of consideration. Shortest path algorithms such as Dijkstra's or Floyd's algorithm are used. Then it generate an $n \times n$ distance matrix M , whose (i, j) entry contains the estimated distance between nodes i and j using above algorithms [2]. In second step, Classical metric MDS is applied to the distance matrix, retaining the first two largest values to construct a relative map that gives location for each node. The locations of nodes may be accurate relative to one another, but the entire map will be arbitrarily rotated and flipped relative to the true node position. The final step transforms the relative map to an absolute map based on the absolute positions of anchors which includes scaling, rotation, and reflection. Main goal is to reduce the errors between the true positions and transform positions of the anchor in MDS map.

2. **RSSI (Received Signal Strength Indicator) based centralized localization technique** uses a theoretical or empirical model to translate signal strength into distance. It consists of three stages. First stage is RF mapping of the network which is obtained by conveying short packets at different power levels through the network and by storing the average RSSI value of the received packets in memory tables. In the second stage all the tuples recorded between the two anchors are processed at the central unit to compensate the non linearity and calibrate the model. In the final stage an optimization problem is solved and provides the position of the nodes. The final result can be obtained by minimizing the function [1].

3. **Localize node based on Simulated Annealing** is based on two stages. In the first stage simulated annealing is used to obtain an estimate of location of the localizable sensor nodes using distance constraints. In the next stage of the algorithm, the error caused by flip ambiguity is eliminated [1].

4. **Semi – Definite Programming (SDP)** algorithm uses the geometric constraints between nodes, represented as linear matrix inequalities (LMIs). Once all the constraints in the network are expressed in this form, the LMIs can be combined to form a single semi definite program [1].

A comparative summary of the four above mentioned centralized techniques is as shown in Table 1. The parameters considered for comparison are principles, techniques, advantages and limitations [1], [2].

Distributed localizations can be categorized into following classes.

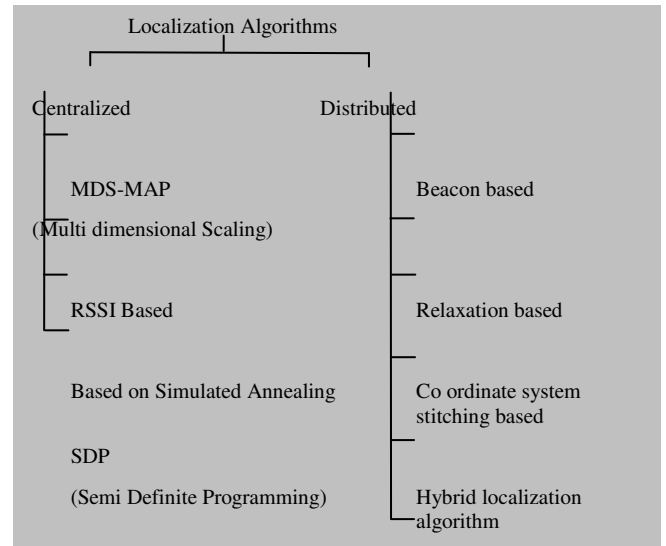


Fig. 2. Elements of Localization [1].

1. *Beacon-based distributed algorithms*: Beacon-based distributed algorithms start with some group of beacons and nodes in the network to obtain a distance measurement to a few beacons, and then use these measurements to determine their own location [1].

2. *Relaxation-based distributed algorithms* use a coarse algorithm to roughly localize nodes in the network [1]. This coarse algorithm is followed by a refinement step, which typically involves each node adjusting its position to approximate the optimal solution.

3. *Coordinate system stitching based distributed algorithm* divides the network in to small overlapping sub regions, each of which creates an optimal local map [1]. Next the scheme merges the local maps into a single global map.

4. *Hybrid localization algorithm* use two different localization techniques such as: multidimensional scaling (MDS) and proximity based map (PDM) or MDS and Ad-hoc Positioning system (APS) to reduce communication and computation cost.

5. *Interferometric ranging based localization* exploits interfering radio waves emitted from two locations at slightly different frequencies to obtain the necessary ranging information for localization.

6. *Error propagation aware localization* uses undesirable wireless environment, such as channel fading and noise corruption. A comparative study of the algorithms for distributed localization is given in Table 2. The parameters considered for comparison are principles, techniques, advantages, and limitations [1], [2], [3].

Summary of Centralized Localization Algorithms

Algorithm	MDS-MAP	RSSI-based technique	Simulated Annealing(SA)	Semi-Definite Progr.(SDP)
Principle	Computes shortest paths between all pairs of nodes in the region of consideration. Algorithm uses the law of cosines and linear algebra to reconstruct the relative position of the points based on the pair wise distances	Localizes nodes through RF attenuation in Electromagnetic waves.	Localize the sensor nodes in a centralized manner, Method is based on neighborhood information of nodes and it works well in a sensor network with medium to high node density.	Based on LMI (Linear matrix inequality)
Technique	Construct the distance matrix for MDS	RF mapping of the network. Creation of the ranging model.	simulated annealing is used to obtain an estimate of location of the localizable sensor nodes using distance constraints	Geometric constraints between nodes are represented as linear matrix inequalities (LMIs). Once all constraints in expressed in this form, the LMIs can be combined to form a single semi definite program, which is solved to produce a bounding region for each node.
Advantages	Does not need anchor or beacon nodes to start with. Works well in situations with low ratios of anchor nodes. High accuracy, Error propagation is low, Low node density, Beacon percentage is low	It is a practical, self-organizing scheme allows addressing any outdoor environments	This algorithm does not propagate error in localization. Gives better accuracy than the semi-definite programming localization.	Its elegance on concise problem formulation, clear model representation, and elegant mathematic solution. High accuracy, Beacon density is medium, error is low.
Limitation/challenging issues	Requires global information of the network and centralized computation. Computation cost is high, Communication cost is high.	Scheme is power consuming.	When the node density is low, the node is flipped & still maintains the correct neighborhood; the proposed algorithm fails to identify the flipped node.	All geometric constraints cannot be expressed as LMIs. Precise range data cannot be conveniently represented. Inability to accommodate precise range data. Computation cost high, Communication cost high.

Summary of Distributed Localization Algorithms	Beacon-based distributed algorithms	Relaxation-based distributed algorithms	Coordinate system stitching based distributed algorithms	Hybrid localization algorithms	Interferometric ranging based localization	Error propagation aware localization
Principle	Estimates distance to reference nodes that may be several hops away.	Nodes estimating their positions with a method such as gradient distance propagation	Localization is originated in a local group of nodes in relative coordinates. By gradually merging such local maps, it achieves entire network localization in global coordinates.	Two orthogonal techniques tailored and combined into a powerful hybrid localization algorithm (RH+).	radio waves emitted from two locations at slightly different frequencies to obtain the necessary ranging information	Integrates the path loss and distance measurement error model
Technique	Distance-vector technique is used to propagate distances from reference node to unknown nodes.	The choice of the nodes is performed using a hop count approximation to distance. The node positions (x_i , y_i) are calculated using polar coordinates	The network is divided into small overlapping sub-regions, creates an optimal local map. The scheme merges the local maps into a single global map.	Multidimensional scaling (MDS) & proximity based map (PDM); Ad-hoc Positioning System (APS).	Taking multiple measurements, it is possible to reconstruct the relative location of the nodes	Anchor nodes broadcast their information (unique ID), global coordinates & position error variance
Advantages	Computation & Communication cost is low.	Fully distributed & concurrent. Operate without beacons	No global resources or communications are needed. Beacon % is low.	Reduce communication & computation cost. Robustness & more accurate.	Gives precise measurements than other common techniques	Precise estimation than other localization schemes
Limitations/ challenging issues	Accuracy is low, Node, Beacon % & Error propagation is high.	Susceptible to local minima, Techniques are quite sensitive to initial starting positions. Local minima problem is worsen at large scales.	Convergence may take some time and that nodes with high mobility may be hard to cover. Low accuracy, High node density, Error propagation is high.	It does not perform well when there are only few anchors.	Requires considerably larger set of measurement which limits their solution to smaller network. Error propagation is significant problem.	Estimation cost is high.

5. OPEN ISSUES

On comparing various algorithms for Centralized and Distributed localization it is observed that the communication cost and computation cost is high in case of MDS-MAP multidimensional scaling and SPD (Semi-Definite Programming). While RSSI based centralized

localization is power consuming. In Beacon based distributed algorithm and Coordinate system stitching based distributed algorithm, though the computation and communication cost is low but error propagation is high, accuracy is low. The Relaxation based distributed algorithm worsens at large scales. Interferometric ranging based localization gives precise measurement than other

techniques but it requires considerably large set of measurement. The estimation cost of Error propagation aware localization is high. We decided to develop new hybrid algorithm that can give maximum possible advantages of both the localization techniques.

6. CONCLUSION

Localization algorithms depend heavily on a variety of factors such as application needs and available physical measurements. No specific algorithm is a clear favorite across all applications. A lot of work still needs to be done to realize practical applications of wireless sensor networks.

REFERENCES

- [1] Amitangshu Pal, Localization Algorithms in Wireless Sensor Networks: Current Approaches and Future Challenges, Network Protocols and algorithms, Microthink Institute, Vol. 2, No. 1.
- [2] Zheng Yang, Localization and Localizability in Sensor and Ad-hoc Networks, Ph.D. Thesis, Hong Kong University of Science & Technology, June 2010.
- [3] Can Basaran. A hybrid localization algorithm for wireless sensor networks, Masters Thesis, Yeditepe University, 2007.
- [4] Azzedine Boukerche, Horacio A. B. F. Oliveira, Eduardo F. Nakamura and FUCAPI Antonio A. F. Loureiro, Secure Localization Algorithms for Wireless Sensor Networks, IEEE Communications Magazine, April 2008.

Securing Mobile Phone Communications

Abubaker Maraicar

*AITTM, Amity University, Noida (UP), India
bakerabu123@yahoo.com*

Abstract: This paper describes how to achieve a good level of security in wireless cellular network. Wireless communication has become an important part of our day to day life. Besides, using cellphones for voice communication, we are accessing the internet, conducting monetary transactions, sending text messages and multimedia messages etc., new services and applications are continued to be added in our cellphones. Therefore, it is essential to provide users with a secure channel of communication.

This paper will give a description of what technology to be adopted for a secured wireless cellular network and limitations of wireless cellular network, their security issues and the different types of attacks. Then the steps taken to overcome the different types of threats in the new generation wireless networks will be provided. Also the security features of WAP being used to access the internet will be discussed. The paper will go over some new security mechanisms that have been proposed by researchers.

1. GENERATION OF CELLULAR NETWORK

A. First Generation (1G)

Cellular Networks have been around since the 1980s and each year their subscribers increase at a very fast rate. First generation (1G) networks were the first cellular networks introduced in the 1980s. They were only capable of transmitting voice at speeds of about 9.6 kbps max. In the US the system was known as Advanced Mobile Phone System (AMPS) and in Europe the Nordic Mobile Telephony (NMT). Both these technologies used analog modulation to transmit data as a continuously varying waveform. 1G system had some limitations such as no support for encryption, poor sound quality and inefficient use of the spectrum due to their analog nature.

B. Second Generation (2G & 2.5G)

Second generation (2G) cellular networks also known as personal communication services (PCS) introduced the concept of digital modulation meaning that voice is converted into digital code, and then into analog (radio) signals. Being digital, they overcame certain limitations of 1G system. Various 2G technologies have been deployed around the world. Code Division Multiple Access (CDMA), North American Time Division Multiple Access (NA-TDMA) and Digital AMPS (D-AMPS) have been deployed in the US, whereas, Global System for Mobile

communication (GSM) has been deployed in Europe and USA and Personal Digital Cellular (PDC) has been deployed in Japan.

Some of the data services which are part of the 2G and 2.5G are

- *Short Messaging Service (SMS):* Transfer of messages between cell phones. Large messages are truncated and sent as multiple messages.
- *High-Speed Circuit-Switched Data (HSCSD):* This was the first attempt at providing data at high speeds data over GSM, with speeds of up to 115 kbps. This technique cannot support large bursts of data. HSCSD was not widely implemented and GPRS became a more popular technique.
- *General Packet Radio Service (GPRS):* This technique can support large burst data transfers. In order to support this two new elements have to be added to existing networks. Service GPRS support node (SGSN) for security mobility and access control and Gateway GPRS support node (GGSN) in order to connect to external packet switched networks.
- *Enhanced Data Rates for GSM Evolution (EDGE):* The standard GSM uses GMSK modulation. Edge uses 8-PSK modulation. GPRS and EDGE combined provide data rates of up to 384 kbps.
- *Cellular Digital Packet Data (CDPD):* CDPD is a packet based data service. CDPD is able to detect idle voice channels and uses them to transfer data traffic without affecting voice communications.

C. Third Generation (3G)

The Third generation (3G) standard is currently being pushed as the next global standard for cellular communications. It will provide services such as fast Internet surfing, advanced value added services and video telephony. Deployments of this technology have already begun and several countries like Austria, Denmark, South Korea and Japan have adopted the 3G network architecture. There are three main technologies that are being applied. In the US CDMA2000, in Europe Wideband CDMA (W-CDMA) and in China Time Division-Synchronous Code Division Multiple Access (TD-SCDMA).

3G is the next generation wireless cellular network whose aim is to provide a worldwide standard and a common frequency band for mobile networking. The International Telecommunication Union (ITU) started the process in 1992, the result of this effort was a new network infrastructure called International mobile telecommunications 2000 (*IMT-2000*), with the 2000 signifying that this new technology will be available in 2000, will have data rates of up to 2000 Kbps and will be in the 2000 MHz frequency range. The following is the list of objectives that IMT-2000 aims[04] to receive :

- To make a wide range of services, both voice and data available to users, irrespective of location.
- To provide services over a wide coverage area.
- To provide the best quality of service (*QoS*) possible.
- To extend the number of services provided subject to constraints like radio transmission, *spectrum* efficiency and system economics.
- To accommodate a great variety of mobile stations.
- To admit the provision of service by more than one network in any area of coverage.
- To provide an *open architecture* which will permit the easy introduction of technology advancements as well as different applications.
- To provide a modular structure which will allow the system to start from small and simple configuration and grow as needed, both in size and complexity within practical limits.

D. Forth Generation (4G)

Although 3G has not been fully deployed, people have already started talking about the fourth generation (4G) technology. This generation will be designed to have data rates of up to 20Mbps. It will have support for next generation Internet such as *IPv6*, *QoS* and Mobile communication over Internet Protocol (*MoIP*), lower system cost and high capacity and capable of supporting communication in moving vehicles with speed up to 250 km/hr.

2. ISSUES IN CELLULAR NETWORKS

The infrastructure for Cellular Networks is massive, complex with multiple entities coordinating together, such as the IP Internet coordinating with the core network. And therefore it presents a challenge for the network to provide security at every possible communication path.

A. Limitation Issues of Cellular Networks

Compared to Wired Networks, Wireless Cellular Networks have a lot of limitations.

- **Open Wireless Access Medium:** Since the communication is on the wireless channel, there is no physical barrier that can separate an attacker from the network. [1]
- **Limited Bandwidth:** Although wireless *bandwidth* is increasing continuously, because of *channel contention* everyone has to share the medium.
- **System Complexity:** Wireless systems are more complex due to the need to support mobility and making use of the channel effectively. By adding more complexity to systems, potentially new *security vulnerabilities* can be introduced.[2]
- **Limited Power:** Wireless Systems consume a lot of power and therefore have a limited time battery life.
- **Limited Processing Power:** The processors installed on the wireless devices are increasing in power, but still they are not powerful enough to carry out intensive processing.
- **Relatively Unreliable Network Connection:** The wireless medium is an unreliable medium with a high rate of errors compared to a wired network.

B. Security Issues In Cellular Networks

There are several security issues that have to be taken into consideration when deploying a cellular infrastructure. The importance of which has increased with the advent of advanced networks like 3G :-

- **Authentication:** Cellular networks have a large number of subscribers, and each has to be authenticated to ensure the right people are using the network. Since the purpose of 3G is to enable people to communicate from anywhere in the world, the issue of cross region and cross provider authentication becomes an issue.
- **Integrity:** With services such as SMS, chat and file transfer it is important that the data arrives without any modifications.
- **Confidentiality:** With the increased use of cellular phones in sensitive communication, there is a need for a secure channel in order to transmit information.
- **Access Control:** The Cellular device may have files that need to have restricted access to them. The device might access a database where some sort of role based access control is necessary.[5]
- **Operating Systems In Mobile Devices:** Cellular Phones have evolved from low processing power, ad-hoc supervisors to high power processors and full fledged operating systems. Some phones may use a *Java* Based system, others use *Microsoft Windows CE* and have the same capabilities as a desktop computer. Issues may

arise in the OS which might open security holes that can be exploited.

- **Web Services:** A Web Service is a component that provides functionality accessible through the web using the standard *HTTP* Protocol. This opens the cellular device to variety of security issues such as *viruses*, *buffer overflows*, *denial of service* attacks etc. [6]
- **Location Detection:** The actual location of a cellular device needs to be kept hidden for reasons of privacy of the user. With the move to IP based networks, the issue arises that a user may be associated with an access point and therefore their location might be compromised.
- **Viruses and Malware:** With increased functionality provided in cellular systems, problems prevalent in larger systems such as viruses and *malware* arise. The first virus that appeared on cellular devices was Liberty. An affected device can also be used to attack the cellular network infrastructure by becoming part of a large scale denial of service attack.
- **Downloaded Contents:** Spyware or Adware might be downloaded causing security issues. Another problem is that of digital rights management. Users might download unauthorized copies of music, videos, wallpapers and games.
- **Device Security:** If a device is lost or stolen, it needs to be protected from unauthorized use so that potential sensitive information such as emails, documents, phone numbers etc. cannot be accessed.

C. Types of Attacks

Due to the massive architecture of a cellular network, there are a variety of attacks that the infrastructure is open to :-

- **Denial Of Service (DOS):** This is probably the most potent attack that can bring down the entire network infrastructure. This is caused by sending excessive data to the network, more than the network can handle, resulting in users being unable to access network resources.
- **Distributed Denial Of Service (DDOS):** It might be difficult to launch a large scale DOS attack from a single host. A number of hosts can be used to launch an attack.
- **Channel Jamming:** Channel jamming is a technique used by attackers to jam the wireless channel and therefore deny access to any legitimate users in the network.
- **Unauthorized Access:** If a proper method of authentication is not deployed then an attacker can gain free access to a network and then can use it for services that he might not be authorized for.

- **Eavesdropping:** If the traffic on the wireless link is not encrypted then an attacker can eavesdrop and intercept sensitive communication such as confidential calls, sensitive documents etc.
- **Message Forgery:** If the communication channel is not secure, then an attacker can intercept messages in both directions and change the content without the users ever knowing.
- **Message Replay:** Even if communication channel is secure, an attacker can intercept an encrypted message and then replay it back at a later time and the user might not know that the packet received is not the right one.
- **Man In The Middle Attack:** An attacker can sit in between a cell phone and an access station and intercept messages in between them and change them.
- **Session Hijacking:** A malicious user can hijack an already established session and can act as a legitimate base station.

3. SECURITY MECHANISMS AND ARCHITECTURE IN 3G

A. Security Mechanisms

This survey paper will not delve into the security features of different 3G architectures. Since the underlying technology is the same, security features of one architecture are applicable to others as well. 3G - UMTS, the most popular of the architectures builds upon the security features of 2G systems so that some of the robust features of 2G systems are retained. The aim of the 3G security architecture is to improve on the security of 2G systems. Any holes present in the 2G systems are to be addressed and fixed. Also, since many new services have been added to 3G systems, the security architecture needs to provide security for these services.

B. 3G Security Architecture

There are five different sets of features that are part of the architecture:

- **Network Access Security:** This feature enables users to securely access services provided by the 3G network. This feature is responsible for providing identity confidentiality, authentication of users, confidentiality, integrity and mobile equipment authentication. User Identity confidentiality is obtained by using a temporary identity called the International Mobile User Identity. Authentication is achieved using a challenge response method using a secret key. Confidentiality is obtained by means of a secret *Cipher Key* (CK) which is exchanged as part of the *Authentication and Key Agreement* Process (AKA). Integrity is provided using

an integrity algorithm and an integrity key (IK). Equipment identification is achieved using the International Mobile Equipment Identifier (IMEI).

- **Network Domain Security:** This feature enables nodes in the provider domain to securely exchange signaling data, and prevent attacks on the wired network.
- **User Domain Security:** This feature enables a user to securely connect to mobile stations.
- **Application Security:** This feature enables applications in the user domain and the provider domain to securely exchange messages.
- **Visibility and Configurability of Security:** This feature allows users to enquire what security features are available.

4. MAJOR SECURITY MECHANISMS AND

ALGORITHMS

A. UMTS Authentication and Key Agreement

The UMTS Authentication and Key Agreement (UMTS AKA) mechanism is responsible for providing authentication and key agreement using the challenge/response mechanism. Challenge/Response is a mechanism where one entity in the network proves to another entity that it knows the password without revealing it. There are several instances when this protocol is invoked. When the user first registers with the network, when the network receives a service request, when a location update is sent, on an attach/detach request and on connection reestablishment. The current recommendation by 3GPP for AKA algorithms is *MILENAGE*. *MILENAGE* is based on the popular shared secret key algorithm called AES or Rijndael. Readers interested in the AES algorithm are encouraged to look at [Imai06]. AKA provides mutual authentication for the user and the network. Also, the user and the network agree upon a cipher key (CK) and an integrity key (IK) which are used until their time expires.

B. Signaling Data Integrity Mechanism

Control Signaling Communication between the mobile station and the network is sensitive and therefore its integrity must be protected. This is done using the *UMTS Integrity Algorithm* (UIA) which is implemented both in the mobile station and the RNC. This is known as the f9 algorithm. Figure 1 [01] shows how this algorithm is applied. First, the f9 algorithm in the user equipment calculates a 32 bit MAC-I for data integrity using the signaling message as an input parameter. This, along with the original signal message is sent to the RNC, where the XMAC-I is calculated and then compared to the MAC-I. If both are same, then we know that the integrity of the message has not been compromised.

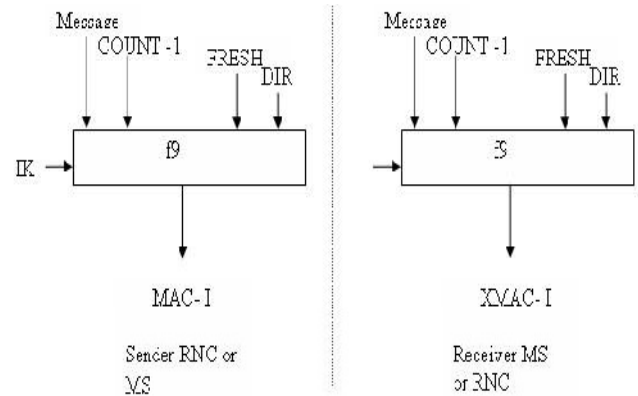


Fig. 1. Signaling Data Integrity Mechanism

C. Air Interface Confidentiality Mechanism

The confidentiality algorithm is known as f8 and it operates on the signaling data as well as the user data. Figure 2 [Imai06] shows how this algorithm is applied. The user's device uses a Cipher Key CK and some other information and calculates an output bit stream. Then this output stream is xored bit by bit with the data stream to generate a cipher stream. This stream is then transmitted to the RNC, where the RNC uses the same CK and input as the user's device and the f8 algorithm to calculate the output stream. This is then xored with the cipher stream to get the original data stream.

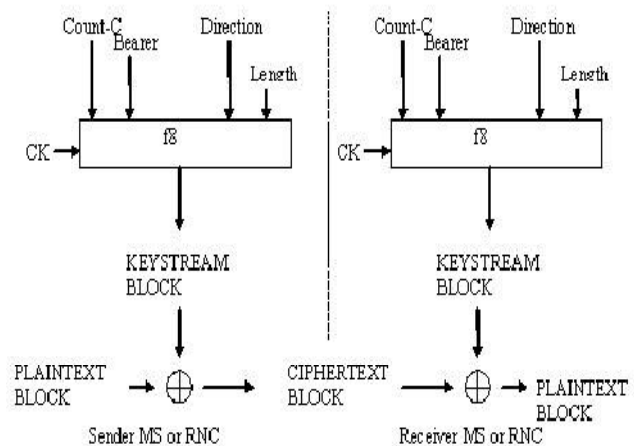


Fig. 2. Air Interface Confidentiality Mechanism

For more information on the inputs to the f8 and f9 algorithms, please refer to [03].

D. KASUMI Block Cipher Mechanism

A block cipher known as the *KASUMI* cipher is central to both the f9 and the f8 algorithm. This cipher is based on the feistel structure using 64 bit data blocks and a 128 bit key.

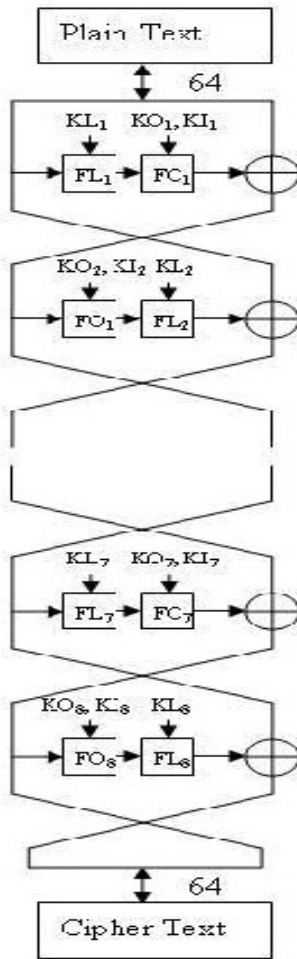


Fig. 3. KASUMI Block Cipher

It has eight rounds of processing, with the plain text (can be any form of data) as input to the first round and the cipher text the result after the last round. An encryption key is used to generate round keys (KL_i, KO_i, KI_i) for each round. Each round calculates a separate function since the round keys are different. The same algorithm is used for encryption and decryption. The KASUMI cipher is based on the *MISTY1* cipher which was chosen by 3GPP due to its proven security against many advanced cipher breaking techniques. It has been optimized for hardware implementation which is important concerning the hardware constraints of cellular devices, such as limited power and limited memory. As shown in the Figure-3 [04], the function f consists of sub functions FL_i and FO_i . FL is a simple function consisting of shifts and logical operations. The FO function is much more complicated and is itself based on the *fiestel* structure and consists of three rounds[04].

E. Wireless Application Protocol (WAP)

Since one of the most important services provided by 3G systems is access to the Internet, it is important to

understand the security mechanisms of the protocol used to access the Internet. WAP is an open specification which enables mobile users to access the Internet. This protocol is independent of the underlying network e.g. *WCDMA*, *CMDA 2000* etc and also independent of the underlying operating system e.g. *Windows CE*, *PALM OS* etc.

The first generation is known as *WAP1* which was released in 1998. WAP1 assumes that the mobile devices are low on power and other resources. And therefore the devices can be simple while sharing the security responsibilities with the gateway devices. The second generation is known as *WAP2* and was released in 2002. WAP2 assumes that the mobile devices are powerful and can therefore directly communicate with the servers. Figure 4 and Figure 5 show the protocol stack for WAP1 and WAP2 respectively.

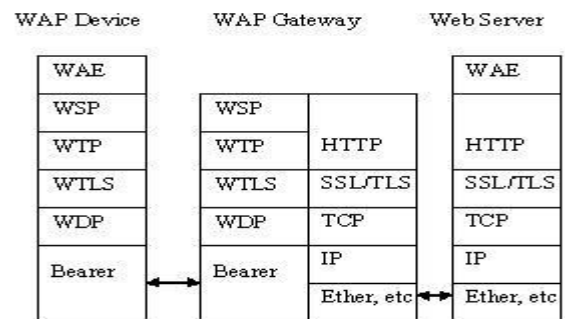


Fig. 4

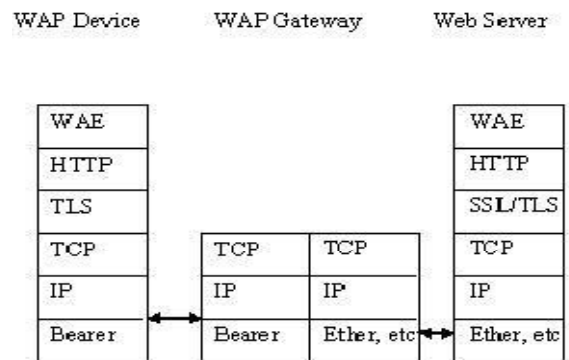


Fig. 5

A brief description of each layer is as follows :-

- **Wireless Application Environment (WAE):** This provides an environment for running web applications or other WAP applications.
- **Wireless Session Protocol (WSP):** This is similar to the HTTP protocol and provides data transmissions with small sizes so that WAP1 clients can process the data with less complexity.

- **Wireless Transaction Protocol (WTP):** This is responsible for providing reliability.
- **Wireless Transport Layer Security (WTLS):** This is responsible for providing security features such as authentication, confidentiality, integrity etc. between a WAP1 client and the WAP gateway.
- **Wireless Datagram Protocol (WDP):** This provides the underlying transport service.
- **Hypertext Transfer Protocol (HTTP):** A standard protocol used to transmit web pages.
- **Transport Layer Security (TLS):** This layer provides security features such as authentication, confidentiality, integrity etc. In WAP1, this is between the WAP1 gateway and the server. In WAP2 this is between the WAP2 client and the server.
- **Transport Control Protocol (TCP):** Standard transport protocol used to provide reliability over IP.
- **Internet Protocol (IP):** Protocol used to route data in a network.
- **Bearer Protocol:** This is the lowest level protocol and can be any wireless technique such as GSM, CDMA etc.

5. ADDITIONAL SECURITY MECHANISM

A. A New Authentication Scheme with Anonymity For Wireless Networks: When a mobile user is roaming, it is necessary to provide anonymity to the users so that malicious parties are unable to associate the user with a particular session. The most basic method to provide anonymity is to have a temporary identity (TID) instead of the real id of the user. There are several issues to consider when designing a security protocol for cellular networks. One, they have low computational power which means that algorithms that require high processing power are not suitable. Second, the error rate of messages increase on wireless networks as compared to cellular networks. Therefore, any mechanism that is designed should minimize message sizes and the number of messages in order to reduce the error rate. The author of [11] specifies an authentication scheme which use public key cryptosystems, hash functions and smart cards and makes use of a temporary key and a temporary certificate. They also show the performance of their algorithms with other algorithms, which clearly show that the proposed method is practical and efficient compared to other algorithms.

B. Manual Authentication For Wireless Devices: This is a technique used by devices to authenticate one another by manually transferring data between the devices. This means that the users will enter some information using some form of input (e.g. keypad). Underneath they employ MAC

algorithms for authentication. Although the scheme that is proposed is secure, its usability depends upon how many numbers (or alphabets) the users have to input [10].

C. Elliptic Curve Cryptography For Wireless Security: Elliptic Curve Cryptography (ECC) is a mechanism which uses points on an elliptic curve to encrypt/decrypt data. It has an advantage over the popular RSA algorithm in that it is much faster. 163-bit ECC provides the same security as a 1024 bit RSA algorithm, and can be anywhere from 5 to 15 times faster depending on the platform. For example, in order to secure a 128 bit AES shared key and 521 - bit ECC provides the same level of security as an 15,360 bit RSA while being about 400 times faster [07].

D. Channel Surfing And Spatial Retreats: Defense against Wireless Denial of Service DOS attacks are one of the most dangerous attacks because they can bring down an entire network. An adversary can either try to fill the buffer in a network device, or can bypass the MAC layer and try to jam the channel. Channel Surfing is a technique where the transmission frequency is changed to one where there is no interface [08].

6. NEW SECURITY MECHANISM PROPOSED BY RESEARCHERS

A. Overview

Traditional mobile security solutions are hardware based. These hardware solutions are, by their very nature, logistically and operationally difficult to deploy. They are not cost effective for use on a large scale. Critically, technology and operational needs often outpace the development of hardware solutions. Therefore hardware solutions will always have a finite period of effectiveness. A software (Secure Application) solution proposed by researchers which can offer the following tremendous benefits to the end user :-

- End user no longer needs bulky and expensive hardware which is difficult to deploy on a large scale.
- The Secure Application Software can be easily downloaded onto existing Mobile Devices. This facilitates fast global deployment in a cost effective way.
- Upgrades can be instantly downloaded to the mobile device.
- The Secure Application can be deployed from a network level, in house from a corporate or government perspective or can be managed by a server for a closed loop end user group.
- The secure application must be compatible with CDMA, LTE networks and 3G networks.

- The secure application should be complete integrated communication solution:-
 - i.. Secure Mobile to Mobile communication.
 - ii.. The secure application software can be configured to operate with authorised Desktop telephones securing office to mobile communication.
 - iii.. A dedicated *APN* may be utilised by the end user to monitor and control internal system billing.
 - iv.. Makes secure and encrypted calling as easy as making a standard network call. The functionality mimics the end users current phone for intuitive ease of use.

B. *Essential Features of Secure Application are shown in table-1*

Table-1

Installation	Pre-installed or easily downloaded under secure conditions
Encryption	Unique encoding keys per call encrypting all data prior to transmission
Coverage	Operates seamlessly over data enabled networks
Ease of use	As easy as making a normal call
Functions	Secure voice, text messaging & voicemail

7. HOW THE SECURE APPLICATION WORKS

The mobile device authorised on the secure network must run the Secure Application. The application runs on a broad range of commercially available standard devices (*iphone, Nokia, Symbian, Android, Black Berry, Windows 7*). The Secure Application :-

- Connects to the best available communication network and establishes a secure encrypted and authenticated connection. The handset will then communicate securely with the servers and any numbers of Secure Application enabled handsets.
- Negotiates one-time encryption keys for all communications. Unique keys are produced per phone for each call, both for transmitted and received voice. These keys are subsequently discarded at the end of each call avoiding any key recovery vulnerabilities. This is the highest level of encryption which exceeds military standards.
- Uniquely functions over all networks including low bandwidth networks such as 2G. This is achieved as the Secure Application applies a unique proprietary codec and algorithms that can run at bandwidth of less than 5 Kbits/sec. Network capacity has up to now been a major constraint in the roll out of viable encryption solutions.

The Secure Application runs as a software download in parallel to standard Smartphone operating systems. The appearance of the Secure Application on screen must be designed to mimic the Operating System of the user's current phone or mobile device for familiarity and ease of use. The software can be downloaded and appears as an application on the users Smartphone. When the user wishes to make a Secure Application secure call, or send Secure Application encrypted data they simply click on the Application icon and then utilise the handset as if they were making a regular call.

- All steps will be completed automatically and instantaneously in the background so that an encrypted call is as easy to make as a regular call for the end user.
- The end user will have the option to make a regular unencrypted call or a Secure Application encrypted call on their handset. Secure Application enabled handset can continue to be used for regular unsecured calls or personal use without compromising secure data or secure communications.

A. Security & Encryption Services

The basis of security and encryption is to protect sensitive information and data. Encryption is the conversion of data using an algorithm into a form that cannot be understood or unencrypted by unauthorised users. It is the replacement of useful, understandable data with a seemingly meaningless arrangement of useless data so that it can only be understood by someone who has the correct decoding key or set of keys. This decoding process allows the conversion of the encrypted data back into its original form, so it can be understood

Encryption is now commonly used in protecting information within many kinds of systems and to protect data in transit, for example, data being transferred via network (e.g internet, e-commerce), wireless systems, money transaction, etc.

Encryption by itself, can protect the confidentiality of data or information, but other techniques are still needed to protect the integrity and authenticity of the message. A single weakness in a system design can allow for a successful attack on the information being protected.

When each secure application authorized phone is powered on comes with range of a suitable data network, it will establish a secure connection configured with security gateway and use that connection to register with the security configured server this connection remain established until the phone is powered off or moves out of range of the communication network. This established connection is used for all signaling functions (making calls, accepting calls, terminating calls etc).

B. Mobile Device Management System

To support the increasing utilization of mobile devices across Govt agencies and corporations some companies has developed a proprietary mobile device management solution to, facilitate standalone, central in-house control and management of critical mobile devices and information. This facility allows the use of personal mobile devices (Mobile phone, Tablet, PDA) for business purpose whilst ensuring that sensitive commercial data remains separate and secure on the device within an encrypted “Sandbox” this allows the organization to centrally manage the commercial data through remote lockdown or wipe, upgrade authorisation and access control criteria without any cross-contamination from personal information stored on the device.

The Secure Application creates a restricted environment through what is termed “*Sandboxing*”. Through this the secure application has the capabilities to store secure encrypted data separately from personal unencrypted data on a smart phone. Sandboxing protects secure data from cross-contamination from unencrypted, unsecured data stored on the device or used for personal purposes. This facilitates major cost savings in supporting the use of a single mobile device for employees rather than using multiple devices for personal and business use.

The Secure Application can, through its encryption data stored on the device can facilitate remote lock-down and remote wiping of encrypted data in cases where the unit is lost, stolen or otherwise compromised. Users can also be remotely decommissioned once they leave or organization and all sensitive data can be removed from the device.

The Secure Application also comes with an optional phone tracking function that can monitor the location of mobile devices. These functions also provide a further significant level of protection for mandown alerts or for financial transactions or corporate data stored within employee’s mobile devices.

C. Trojan Detection Software

Trojan detection software is available for both Secure Application and non Secure Application calls. A Trojan detection scanner protects *critical data* by keeping handsets free of malicious software or spyware so data is always safe. Key features of the software are as under :-

- Protected against Trojans, viruses, *malware & spyware*.
- Safe & reliable firewall protection.
- Self updating virus scanning for 24/7 protection.
- Protects both GSM and MSA calls.

- Runs seamlessly in the background of the device. No user input required.

This detection and prevention feature is very important for mobile security. People are now using Mobile devices in the same way they use PCs. They surf the web, email, text message, Facebook, and download. They shop, carry out banking activity and access business related sites from their devices. All of which expose them to the same vulnerabilities they face on their PC. Individuals, corporations and agencies vigilantly protect PCs, the secure application allows them do the same with mobile devices.

Mobile devices will house more and more personal and commercially sensitive information-contacts, emails, text messages and more. A Trojan may be disguised as legitimate software. They have the ability to record information and then send it to a remote server. When a Trojan is installed on a device it has the ability to completely hide itself from the user. The secure application fully protects the device from any of these vulnerabilities.

Antivirus runs seamlessly in the background with no interaction with the end user unless a breach in security or intrusion is detected on the device.

8. CONCLUSION

Cellular Networks are open to attacks such as DOS, channel jamming, message forgery etc. Therefore, it is necessary that security features are provided that prevent such attacks. The 3G security architecture provides features such as authentication, confidentiality, integrity etc. Also, the WAP protocol makes use of network security layers such as *TLS/WTLS/SSL* to provide a secure path for HTTP communication.

Although 3G provides good security features, there are always new security issues that come up and researchers are actively pursuing new and improved solutions for these issues. People have also started looking ahead at how new features of the 4G network infrastructure will affect security and what measures can be taken to add new security features and also improve upon those that have been employed in 3G.

9. FUTURE

Security is an ever growing field. What is secure today may not be secure tomorrow. There will always be malicious users trying to exploit and find new holes in a network. Therefore, we need to look into the future so that we are able to face these security issues before they cause damage.

A. A Look at Security in 4G

4G is the next generation after 3G. Although still 3G has not been fully implemented in the real world, people have

started talking about the features of 4G. Some of the 4G services talked about are incorporating quality of service (QoS) and Mobility. There is also a concept of always best connected which means that the terminal will always select the best possible access available. 4G will also make use of the IPV6 address scheme. This might make it possible for each cell device to have its own IP address. Currently, the problem of security is solved by using multiple layers of encryption of the protocol stack. There are disadvantages in this scheme such as wasted power, wasted energy and a larger transmission delay. In 4G there will be a concept of interlayer security where only one layer will be configured to do encryption on data.

REFERENCES

- [1] Imai, H., "Wireless communications security," Boston: Artech House, 2006
- [2] Yang, H., "Securing A Wireless World," Proceedings of The IEEE v. 94 no. 2 Feb. 2006
<http://www.cs.ucla.edu/~hyang/paper/ProcIEEE05.ps>
- [3] Xenakis, C., "Security In Third Generation Mobile Networks," Computer Communications 27 (2004) pg.638 to 650
<http://www.cs.uakron.edu/~dang/CS655/Spring05/3G.pdf>
- [4] Balderas-Contreras, T., "Security Architecture in UMTS Third Generation Cellular Networks," Coordinación de Ciencias Computacionales INAOE, Reporte Técnico No. CCC-04-002 27 de febrero de 2004
<http://ccc.inaoep.mx/Reportes/CCC-04-002.pdf>
- [5] Fernandez, E., "An overview of the security of wireless networks," Handbook of Wireless LANs, CRC Press (2004)
<http://polaris.cse.fau.edu/~ed/WirelessSecSurv4.pdf>
- [6] Fernandez, E., "Some security issues of wireless systems," Advanced Distributed Systems: 5th International School and Symposium, ISSADS 2005, Guadalajara, exico, January 24-28, 2005, Revised Selected Papers
http://www.cse.fau.edu/%7Eed/Fernandez_ISSADS2005Final.pdf
- [7] Lauter, K., "The Advantages Of Elliptic Curve Cryptography For Wireless Security," Wireless Communications, IEEE Feb 2004 Volume: 11, Issue: 1 On page(s): 62- 67
<http://research.microsoft.com/~klauter/IEEEfinal.pdf>
- [8] Xu, W., "Channel Surfing and Spatial Retreats: Defenses against Wireless Denial Of Service," Proceedings of the 2004 ACM workshop on Wireless security, 2004
http://www.winlab.rutgers.edu/~trappe/Papers/WiDoS_Wise04.pdf
- [9] Carneiro, G., "Cross-Layer Design In 4G Wireless Terminals," IEEE Wireless Communications, 2004
<http://paginas.fe.up.pt/~mricardo/doc/journals/crossLayerDesign.pdf>
- [10] Gehrmann, C., "Manual authentication for wireless devices," RSA Cryptobytes, 2004
<http://www.isg.rhul.ac.uk/~cjm/mafwd4.pdf>
- [11] Zhu, J., "A new authentication scheme with anonymity for wireless environments," Consumer Electronics, IEEE Transactions on Publication Feb 2004 Volume: 50, Issue: 1 page(s): 231- 235
<http://www.csl.mtu.edu/cs6461/www/Reading/Zhu04.pdf>

Security Challenges in Cloud Computing and SAML: A Study

Sarah J. Andrabi¹, Inhas Ashraf², Roohie Naaz Mir³, Shabir A. Sofi⁴

^{1,2}Dept. of Information Technology, NIT Srinagar, Srinagar, India

¹sarahjameeel@yahoo.com; ²inhas_17@yahoo.com

^{3,4}Faculty, Dept. of Information Technology, NIT Srinagar, Srinagar, India

nazz310@yahoo.co.in; shabir@nitsri.net

Abstract: Cloud computing is transforming the way we look at IT and is turning us into remote users by being a method of delivering hosted services over the internet in a cost-effective and fast manner. Currently trust in the security of cloud computing—both in terms of technology and process—are the number one obstacle slowing its growth. The security issues relating to the ownership of Cloud servers and the exploitation of the information by the service provider is one of the aspects while authentication, identity management and single sign-on remain the key challenges after a cloud computing move. To deal with the issue, Security Assertion Markup Language (SAML) enjoys a position in terms of industry acceptance and production federated identity deployments. In this paper we have discussed some of these challenges and also some traditional encryption standards and have analyzed SAML with respect to security issues in cloud.

Index Terms: Cloud Computing, Security Assertion Markup Language (SAML), Identity Provider (IdP), Service Provider (SP), Single Sign-On (SSO), Key Concepts of SAML, Identity Based Encryption (IBE), Identity Based Signature, Data loss prevention (DLP), Software as a service (SaaS)

1. INTRODUCTION

Cloud computing is an umbrella term [1] and is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation [2]. It is an emerging computing model by which users can gain access to their applications from anywhere, through any connected device. A user-centric interface makes the cloud infrastructure supporting the applications transparent to users. The applications reside in massively scalable data centers where computational resources can be dynamically provisioned and shared to achieve significant economies of scale. [3]

Magnusson, referred to cloud computing as the “fourth wave” of computing following mainframes, client-servers and the Internet, and is definitely a cloud booster. “The more

people use the cloud, the more they like it,” he says. His assertion can be applied not only to the end users but also the service providers and businesses that use Cloud infrastructure. There is no necessity for a high processing or computing power at the user end and service providers can increase the required infrastructure when needed in a cost-effective and efficient way.

The cloud model provides a user experience by which hardware, software and network resources are optimally leveraged to provide innovative services over the Web, and servers are provisioned in accordance with the logical needs of the service using advanced, automated tools. The cloud enables the service creators, program administrators and others to use these services via a Web-based interface that abstracts away the complexity of the underlying dynamic infrastructure. [3]

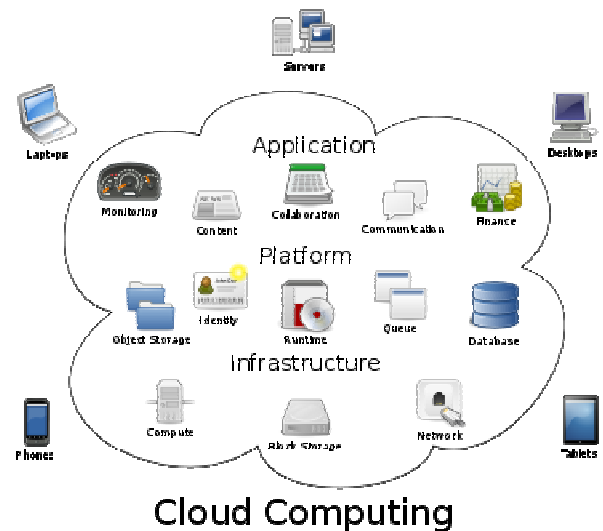


Fig. 1. Cloud Computing Logical Diagram [2]

Cloud computing is where an application doesn't access resources it requires directly; rather it accesses them through

something like a service. So instead of talking to a specific hard drive for storage and a specific CPU for computation, etc. it talks to some service that provides these resources. The service then maps any requests for resources to its physical resources, in order to provide for the application. [5] The services provided are from a pool of large number of resources and can be dynamically allotted as need be.

2. SECURITY CHALLENGES OF CLOUD

A lot of companies are waiting and watching whether cloud computing will prove to be and deliver what it claims to deliver. What are the “security” concerns that have warranted this cautious approach towards Cloud computing? Numerous studies, for example IDC’s 2008 Cloud Services User Survey of IT executives, cite security as the number one challenge for cloud users. [6]

Correct security controls should be implemented according to asset, threat, and vulnerability risk assessment matrices [7]. Cloud security concerns can be grouped into any number of dimensions. The major ones are:

- *Physical Security and Third Party Control*— The service providers must ensure that the physical machines and the data on it are well secured and access to it is highly restricted. This type of a security challenge exists for all type of systems but in cloud computing it becomes even more important to protect data. In cloud computing the threat of company proprietary information is not only from intruders but also from the cloud providers, who can easily misuse or abuse the information they are storing on their servers. Thus access to the data must only be restricted but also documented.

There is also a potential lack of control and transparency when a third party holds the data [10]. Part of the hype of cloud computing is that the cloud can be implemented independent, but in reality regulatory compliance requires transparency into the cloud. All this is prompting some companies to build private clouds to avoid these issues and yet retain some of the advantages of cloud computing. [10]

- *Information Integrity*— How can the user be sure that the information being provided is correct or up-to-date? How can the cloud user be sure that if a certain piece of information that was deleted has actually been deleted? These queries again question the reliability and integrity of the cloud provider and how safe our data actually is.
- *Transitive nature*— Another possible concern is that the contracted cloud provider might itself use subcontractors, over whom the cloud user has even less control, and who also must be trusted. One example is the online storage service called The Linkup, which in

turn used an online storage company called Nirvanix. The Linkup shutdown after losing sizeable amounts of customer data, which some say was the fault of Nirvanix [11].

- *Identity Management*— Every cloud based enterprise will have a mechanism for controlling access to its information and resources. To integrate applications from different application domains, the enterprise will use access management to enable its end users to access applications without re-authentication [12]. This may work well for applications within the data center but since most cloud computing service providers are typically in external data centers and located within different domain, thus requiring a new Single Sign-on [12]. Even if somehow this problem is solved, but since the cloud is outside the firewall, protection of the infrastructure used to connect and interact with the cloud is further complicated [10], and the data thus becomes more vulnerable to attack by intruders.
- *Heterogeneity*—Often times an enterprise can have multiple service providers teaming up to provide various services but having different security approaches and mechanisms to implement their security policies. This can result in a lack of trust framework to handle dynamic interactions between the service providers [12] and discrepancy between the mechanisms to handle the data, resulting in mismanagement and also faulty transactions of data.
- *Access to Data*—Lack of well define constraints on OS services. For example, authorization to define access to well-defined parts of the file system in a multitenant cloud service [22].
- *Uptime*— As with the Traditional Security concerns, cloud providers argue that their server uptime compares well with the availability of the cloud user’s own data centers. Besides just services and applications being down, this includes the concern that a third-party cloud would not scale well enough to handle certain applications [10]. SAP’s CEO, Leo Apotheker said: “There are certain things that you cannot run in the cloud because the cloud would collapse...Don’t believe that any utility company is going to run its billing for 50 million consumers in the cloud.”

3. SECURITY ASSERTION MARKUP LANGUAGE (SAML)

Security Assertion Markup Language (SAML) is an XML based standard for web browser single sign-on (SSO) [14] and defines a framework for exchanging security information between online business partners [15]. It is defined by the OASIS (Organization for the Advancement of Structured Information Standards) Security Services Technical Committee.

SAML is different from other security systems due to its approach of expressing assertions about a subject that other applications within a network can trust [16]. What does this mean? To understand the answer, you need to know the following two concepts used within SAML:

- *Identity Provider (IdP)*: The system, or administrative domain, that asserts information about a subject. For instance, the Identity Provider asserts that this user has been authenticated and has given associated attributes. In SAML, Identity Providers are also known as SAML authorities and Asserting Parties. [15]
- *Service Provider (SP)*: The system, or administrative domain, that relies on information supplied to it by the Identity Provider [15]. A service provider is a website that hosts applications [18]. It is up to the Service Provider as to whether it trusts the assertions provided to it. SAML defines a number of mechanisms that enable the Service Provider to trust the assertions provided to it. It should be noted that although a Service Provider can trust the provided assertions provided, local access policy defines whether the subject may access local resources. Service Providers are also known as Relying Parties – due to the fact that they “rely” on information provided by an Identity Provider (Asserting Party). [15]

Now we can redefine SAML on the basis of identity and service providers as, Security Assertion Markup Language is an XML-based open standard for exchanging authentication and authorization data between security domains, that is, between an identity provider and a service provider [16] and the format of these XML messages and assertions is defined in a pair of SAML XML schemas [17]. All of the requests and responses are transmitted within a SOAP (Simple Object Access Protocol) envelope via HTTP (Hypertext Transfer Protocol) [17].

The biggest advantage or use of SAML is for Single Sign-On services. It provides a framework to implement platform-neutral, secure and scalable SSO solution. Single sign-on (SSO) is a property of access control of multiple related, but independent software systems. With this property a user logs in once and gains access to all systems without being prompted to log in again at each of them [16].

A. Key Concepts of SAML

The key concepts that build of the framework of SAML include:

- Assertions
- Protocol
- Bindings
- Profiles

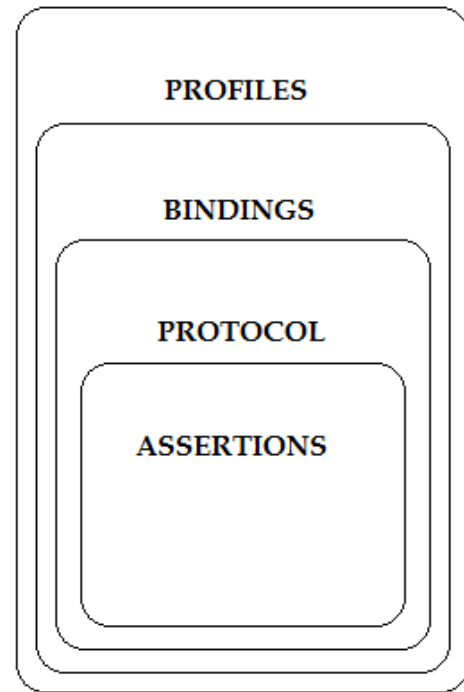


Fig. 2. Key Concepts of SAML

1) *Assertions*: Lying at the core of SAML, assertions contain a packet of security information [16] and are used by the asserting party to communicate the authentication, attributes and entitlement information for a given subject. Assertions are created by Identity Providers [12] and transferred to Service Providers. Loosely speaking, a relying party interprets an assertion as follows: Assertion A was issued at time t by issuer R regarding subject S provided conditions C are valid [16].

2) *Protocol*: Protocols are request and response elements for packaging assertions [12]. A SAML request can either ask for a specific known assertion or make authentication, attribute, and authorization decision queries, with the SAML response providing back the requested assertions [17]. The most important type of SAML protocol request is called a query. A service provider makes a query directly to an identity provider over a secure back channel. Thus query messages are typically bound to SOAP [16].

3) *Bindings*: The lower-level communication or messaging protocols (such as HTTP or SOAP) that the SAML protocols can be transported over are defined by Bindings [15].

4) *Profiles*: SAML Protocols and Bindings, together with the structure of Assertions, can be combined together to create a Profile [15]. They are technical descriptions of particular flows of assertions and protocol messages derived from use cases [17].

4. TECHNIQUES FOR DEALING WITH THE SECURITY CHALLENGES OF CLOUD COMPUTING

It is but expected that everyone will have security as their number one fear when a new technology is introduced to the market. But when you really look deeper into the problem, you'll realize that these security breaches happen because the organization allows them to happen. Cyber criminals often look at cloud computing loopholes and attach those which have loose controls in place [21]. Methods in which security can be implemented in Cloud computing are:

A) Encryption in the Cloud:

There are several encryption techniques that can be deployed to protect sensitive data stored in a Cloud application. Every data protection strategy should consider both data in transit as well as data at rest. And, unfortunately, while most cloud service providers support encryption for data in transit, few offer support for data at rest [23]. A few encryption solutions that can be implemented:

1) Identity Based Encryption (IBE): IBE is a form of public key cryptography in which a third party server uses a simple identifier, such as an e-mail address, to generate a public key that can be used for encrypting and decrypting electronic messages. Compared with typical public key cryptography, this greatly reduces the complexity of the encryption process for both users and administrators [25]. An added advantage is that a message recipient doesn't need advance preparation or specialized software to read the communication [25].

The success of IBE depends upon the third party IBE server that generates private keys. The only information this server stores permanently is a secret master key- a large random number that is exclusive to the security domain [25].

The server uses this key to create a common set of public key parameters that are given to each user who installs the IBE software, and recipient's private keys are required. When a sender creates an encrypted message, the IBE software on his system uses three parameters to generate the public key for the message: a starting value, the current week number and the recipient's identity [25]. A user who receives an IBE encrypted e-mail message but has not used the process before can request- upon authentication- a private key that allows him to decrypt all e-mails encrypted using his e-mail address as the public key [25].

2) Identity Based Signature: An identity based signature scheme is deterministic if the signature on a message by the same user is always the same. The framework of identity based signature scheme consists of algorithms described below:

a) *Setup*: The private key generator (PKG) provides the security parameter as the input to this algorithm, generates the systems parameters and the master private key. [25]

b) *Extract*: The user provides his identity ID to the PKG. the PKG runs this algorithm with identity ID, parameters and master private key as the input and obtain the private key D. the private key D is sent to user through a secure channel. [25]

c) *Sign*: For generating a signature on a message m, the user provides his identity ID, his private key D, parameters and the message m as input. This algorithm generates a valid signature on message m by the user.

d) *Verify*: This algorithm on input a signature on message m by the user with Identity ID, parameters, checks whether signature is valid on message m by ID. [25]

B. Data protection and administrative policies:

Migration to cloud raises many security issues particularly in reference to data protection. Consequently, a thorough examination into and understanding of cloud computing and data protection options is essential for every enterprise [23]. Also, the company will have to be workload-focused instead of cloud-focused. When moving to the clouds, the organization must take into consideration each workload so that it will be able to enforce a security program which is focused on the workload with a possibility to implement non-traditional security measures [21].

Data loss prevention (DLP) tools can help control migration of data to the cloud and also find sensitive data leaked to the cloud [23]. One of the most useful ways to use DLP for cloud computing is to monitor, and even block, data migrations to the cloud from your traditional infrastructure [24]. The vast majority of cloud computing services rely on HTTP as their main out-of-the box communications protocol (albeit often through custom APIs). Thus, if you monitor HTTP (and HTTPS), you'll catch many potential data migrations across the spectrum of cloud service models [24].

A major advantage of cloud computing is that it is capable of virtualization and because of this advantage an organization must have a management process for its storage image implemented [21]. This will guarantee that the required images are made available when needed. The images must also be appropriately managed and identified so that image sprawl will be avoided [21].

5. SAML USE CASES—IMPLEMENTING SECURITY IN CLOUD

Cloud computing is about gracefully losing control while maintaining accountability even if the operational

responsibility falls upon one or more third parties [12][19][20]. Despite the various encryption techniques that are used to secure data transmissions in the cloud, there is still a need to implement greater security without affecting performance. This is where SAML comes into play.

A use case is a list of steps typically defining interactions between a role and a system, to achieve a goal. In case of SAML the use cases are mostly related to sign-on problems and inter-DNS authentication, and website security with implications for implementing security in the Cloud.

The various SAML use cases are:

A. Single Sign-On

This is the original use case as supported in SAML 1.0 and 1.1. Most existing Single-Sign On products use browser cookies to maintain state so that re-authentication is not required [15]. Browser cookies are not transferred between DNS domains. So, if you obtain a cookie from one site, then that cookie will not be sent in any HTTP messages to another site [15]. Now while using SAML, if a user has logged on to some site, and latter decides to logon to another site, then instead of asking the user to re-authenticate himself, what happens is that second site asks previous site where the user had already authenticated whether the user has been authenticated or not. That site then sends a SAML assertion which will indicate that the user has already been authenticated; as such the second site allows the user to proceed without requiring the user to login again. Single Sign-on also establishes Identity federation, that is, a SAML bridge that allows users to use Identity Providers to login into SAML enabled Software as a service (SaaS) endpoints using SAML assertion. SaaS services are configured to accept federated authentication using SAML from partner Identity Providers [12] [19] [20], as explained in the above example.

B. Delegated Authentication

Using delegated authentication, the SaaS service provider does not use SAML assertions but instead uses an external Web service to validate user credentials. When a user attempts to login, the platform checks the user's profile to see if they are enabled for SSO. If so, it makes a Web services call to the endpoint specified for the organization, asking it to validate the username and password [12] [19] [20].

C. Trust Domains

In this solution a user can have different credentials in each application or cloud service. When these applications and cloud services are in a chained trust domain, the SAML identity provider can reconcile different identities allowing

users to access different applications using their appropriate credentials [12][19][20].

D. Distributed Transaction Service

This case represents the case when one site can request another site for a user's profile using a SAML assertion. Say a user orders something from one site and then decides to make another purchase from another site. Now the second site can send a SAML assertion to the previous site, to send the profile of the user, which then sends the user's profile in another SAML assertion statement.

E. Token Translation

Token translation using SAML is now quite an established way to allow applications in one security domain to communicate with applications in another security domain, on behalf of a user whose identity does not have to also flow with the data [26]. In this solution a client has authenticated with Identity Provider. When the client tries to access a SaaS service a Security Token Service converts the security token that was used locally into standard SAML security token containing the user's identity. This token is shared with SaaS. The SaaS provider validates incoming security tokens and generates a new local token for consumption by other applications [12] [19] [20].

F. Website Security

SAML Assertions can be conveyed by means other than the SAML Request/Response protocols or Profiles defined by the SAML specification set. One example of this is the use of SAML by Web Services Security (WSS) [15]. WSS defines a set of SOAP header extensions for end-to-end messaging security [12] [19] [20]. The primary services provided WSS by are Authentication, Data Integrity and Confidentiality. The SAML Assertions as defined by WSS usually play a role in the protection of the message they are carried in, typically they contain a key used for digital signatures; and the SAML assertions typically pertain to the identity of the sender [15].

6. CONCLUSION

As is implied by today's technology scenario, security in cloud remains a big constraint to its widespread adoption in industry. The advantages that cloud computing offers have led to an extensive research in finding ways to overcome one of the biggest shortcoming i.e. security. SAML offers a number of ways for tackling some of the security problems of Cloud computing, though SAML mainly helps in areas related to user authentication and authorization. The various ways SAML deals with these problems without compromising on user efficiency and productivity, on the contrary it is increasing them. Apart from these areas, an essential part of all security mechanisms, encryption, is also

essential for making the data unintelligible to the intruder or even the service provider if they choose to use the data illegally and without permission. There are several ways in which enciphering can be achieved in Cloud; we have the usual algorithms of Triple DES, AES, RSA etc. and other identity based algorithms, that apart from implementing authorization, serve the usual purpose of encryption as well.

REFERENCES

- [1] An Introduction to Cloud Computing, UKOLN: Supporting the Cultural Heritage Sector
- [2] Cloud Computing, Wikipedia
- [3] Seeding the Clouds: Key Infrastructure Elements for Cloud Computing, IBM February 2009
- [4] Understanding Cloud Computing -V.Venkatesa Kumar—
<http://www.code2cloud.com/the-cloud-computing>
- [5] <http://stackoverflow.com/questions/1067987/what-is-the-difference-between-cloud-computing-and-grid-computing>
- [6] IT Cloud Services User Survey, pt.2: Top Benefits & Challenges. <http://blogs.idc.com/ie/?p=210>.
- [7] “Cloud Computing Security Policies you must know”
<http://cloudcomputingsec.com/268/4-cloud-computing-security-policies-you-must-know.html>.
- [8] “Gartner: Seven cloud-computing security risks”
<http://www.infoworld.com/d/security-central/gartner-seven-cloud-computing-security-risks-853>.
- [9] “Security Guidance for Critical Areas of focus in Cloud computing”
<https://cloudsecurityalliance.org/research/projects/security-guidance-for-critical-areas-of-focus-in-cloud-computing/>
- [10] *Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control*, Richard Chow, Philippe Golle, Markus Jakobsson, Ryusuke Masuoka, Jesus Molina, Elaine Shi, Jessica Staddon
- [11] *Loss of customer data spurs closure of online storage service The Linkup*.
<http://www.networkworld.com/news/2008/081108-linkup-failure.html?page=1>.
- [12] <https://cloudsecurityalliance.org/csaguide.pdf>
- [13] CLOIDIFIN:http://community.zdnet.co.uk/blog/0,1000000567,2000625196b,00.htm?new_comment.
- [14] <http://www.onelogin.com/saml/>
- [15] Security Assertion Markup Language (SAML) 2.0 Technical Overview http://www.oasis-open.org/committees/documents.php?wg_abbrev=security
- [16] *Security Assertion Markup Language*
http://en.wikipedia.org/wiki/Security_Assertion_Markup_Language
- [17] http://it.toolbox.com/wiki/index.php/SAML_architecture
- [18] *About Identity providers and Service Providers*
https://login.salesforce.com/help/doc/en/identity_provider_about.htm
- [19] <http://csrc.nist.gov/publications/drafts/800-144/Draft-SP-800-144-cloud-computing.pdf>
- [20] <http://www.sis.pitt.edu/~jjoshi/courses/IS2620/Spring11/cloud.pdf>
- [21] <http://www.cloudtweaks.com/2012/04/how-to-implement-cloud-computing-security/>
- [22] <http://www.slideshare.net/NRAJRAO/saml-in-cloud>
- [23] <http://searchcloudsecurity.techtarget.com/tutorial/Cloud-computing-and-data-protection-Cloud-computing-encryption-tutorial>
- [24] <http://searchcloudsecurity.techtarget.com/tip/Using-DLP-tools-for-cloud-computing-security>
- [25] Cryptography and Encryption In Cloud Computing, Simarjeet Kaur, VSRD International Journal of CS & IT Vol. 2 (3), 2012
- [26] *Connecting SOA to cloud: Using Token Translation and SAML to link domains together*
<http://www.soatothecloud.com/2009/11/using-token-translation-and-saml-to.html>

Implementation of Digital Signature Algorithm for Improved Performance for Small Data Sets

Vijay Kumar Tiwari¹, Anuraag Awasthi², Ritesh Rastogi³, Anuj Kumar⁴

¹vijayvijay456@gmail.com, ²anuraag_awasthi@hotmail.com, ³rit_ras@hotmail.com, ⁴a.kumaranuj007@gmail.com

Abstract: A digital signature or digital signature scheme is a mathematical scheme for demonstrating the authenticity of a digital message or document. A valid digital signature gives a recipient reason to believe that the message was created by a known sender, and that it was not altered in transit. Digital signatures are commonly used for software distribution, financial transactions, and in other cases where it is important to detect forgery or tampering.

Digital signatures are often used to implement electronic signatures, a broader term that refers to any electronic data that carries the intent of a signature, but not all electronic signatures use digital signatures. In some countries, including the United States, India, and members of the European Union, electronic signatures have legal significance. However, laws concerning electronic signatures do not always make clear whether they are digital cryptographic signatures in the sense used here, leaving the legal definition, and so their importance.

I am trying to implement the concept of Digital Signature in a new and efficient manner. I studied various algorithms that are currently available to solve the problem of digital signature implementation, and after the analysis I found out that more efficient and secure algorithm can be developed by using positive aspects of various algorithms with some enhancements and by doing some good technical implementations in them.

Keywords: cryptography, digital signatures, Morph Digital signature, Processing Standards, public key cryptography, RSA, DSS.

1. INTRODUCTION OF DIGITAL SIGNATURE

A digital signature is a term used for marking or signing an electronic document by a process meant to be analogous to paper signatures, but which makes use of a technology known as private key cryptography. In other words, it's a digital code that can be attached to an electronically transmitted message that uniquely identifies the sender and insures that the document has not been altered in any way since the sender has signed it.

A Digital Signature is a construct which helps achieve non-repudiation of Origin of data by digitally signing the document, the person who signs it assures that he is the author of the document or the message that was signed^[1]. A digital signature can also be used to verify that information has not been altered after it was signed. Digital signatures

rely on certain types of encryption to ensure authentication^[2].

The Information Technology Act of India 2000 provides for the use of Digital Signatures on the documents submitted in electronic form in order to ensure the security and authenticity of the documents. Digital Certificates are issued only through a valid Certification Authority (CA), such as e-Mudhra^[3].

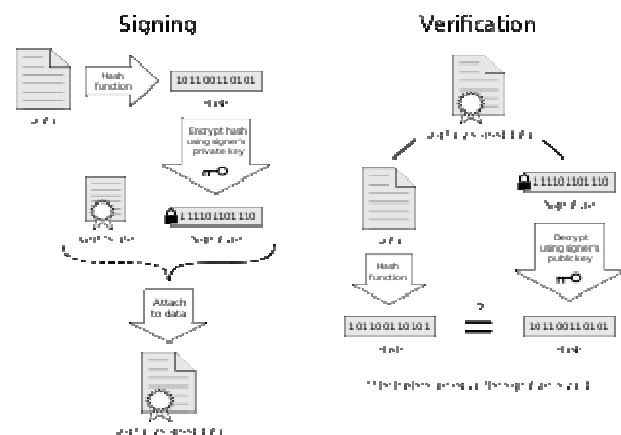


Fig. 1. Process of digital signature

Digital signature is a sort of Cryptography. Cryptography means keeping communications private. It deals with encryption, decryption and authentication. It is a practical art of converting messages or data into a different form, such that no one read them without having access to the 'key'.

Digital signature is a sort of Cryptography. Cryptography means keeping communications private. It deals with encryption, decryption and authentication^[4]. It is a practical art of converting messages or data into a different form, such that no one read them without having access to the 'key'^[5].

There are two types of Cryptography

1. Secret key or Symmetric Cryptography: In this the sender and receiver of a message know and use the same secret key to encrypt the message, and the receiver uses same key to decrypt the message^[6].

- Public key or Asymmetric Cryptography: This cryptography involves two related keys, one of which only the owner knows (the 'private key') and the other which anyone can know (the 'public key').

1.1 Need for Digital Signature

During the "E" revolution, there was a need for authenticating critical transactions especially in the financial World. If Vijay has agreed to transfer ₹x to Ram, then there had to be a way for Bob to be sure that:

- It was Ram who performed the transaction and not someone else impersonating Ram (Authentication)
- The amount agreed by Ram is ₹x (Integrity)
- Shyam could not dispute her statement of transacting ₹x to Ram (Non-Repudiation of Origin)

The most attractive solution to this problem is the digital signature^[7].

1.2 Requirements

Following are the requirements for a digital signature:

- The signature must be a bit pattern that depends on the message of being signed.
- The signature must use some information unique to sender, to prevent both forgery and denial.
- It must be relatively easy to produce the digital signature.
- It must be relatively easy to recognize and verify the digital signature.
- It must be computationally infeasible to forge a digital signature, either by constructing a new message for an existing digital signature or by constructing a fraudulent digital signature for a given message.
- It must be practical to return a copy of the digital signature in storage.

McEliece Based Digital Signature Scheme

McEliece is one of the oldest known public key cryptosystems. It had been proposed back in the 70s. It is based on intractability of respectively factorization and syndrome decoding problem

The McEliece cryptographic scheme is based on error correcting codes. It consists in randomly adding errors to a

codeword (as it would happen in a noisy channel) and uses this as a cipher. The decryption is done exactly as it would be done to correct natural transmission errors^[8]. The security of this scheme simply relies. How to Achieve a McEliece-Based Digital Signature Scheme 159 on the difficulty of decoding a word without any knowledge of the structure of the code? Only the legal user can decode easily using the trap^[9].

Public key: G

Clear text: x belongs to $\text{pow}(F_2, k)$

Cipher text: $y = xG + e$, $wH(e) = t$

Cipher text space: $\text{pow}(F_2, n)$

Shortcomings

- Having such short signatures enables attacks independent on the strength of the trap door function used, which are inherent to the commonly used method of computing a signature by in version of the function.
- The cracking problem for McEliece is the problem of decoding an error correcting code called Syndrome Decoding (SD).
- It generates the Digital Signatures of 81-bits. Hence security factor is only of $\text{pow}(2, 83)$.
- In many cases, code-based cryptosystems like McEliece do not allow practical digital signatures.

2. DSA

A digital signature algorithm is used by a signatory to generate a digital signature on data and by a verifier to verify the authenticity of the signature^[10]. Each signatory has a public and private key. The private key is used in the signature generation process and the public key is used in the signature verification process^[11].

DSA PARAMETERS

The DSA makes use of the following parameters:

- p = a prime modulus
- q = a prime divisor of $p - 1$
- $g = \text{pow}(h, (p-1)/q) \bmod p$, where h is any integer with $1 < h < p - 1$ such that $\text{pow}(h, (p-1)/q) \bmod p > 1$ (g has order $q \bmod p$)
- x = a randomly or pseudo randomly generated integer with $0 < x < q$
- $y = \text{pow}(g, x) \bmod p$

7. k = a randomly or pseudo randomly generated integer with $0 < k < q$

The integer's p , q , and g can be public and can be common to a group of users. A user's private and public keys are x and y , respectively. They are normally fixed for a period of time. Parameters x and k are used for signature generation only, and must be kept secret. Parameter k must be regenerated for each signature.

Limitations

DSA is a highly secure algorithm. The only known way to attack it is to perform a "brute-force" attack on the modulus. This attack can be easily defeated by simply increasing the key size. However, this approach can lead to a number of serious problems:

3. THE KOREAN CERTIFICATE BASED DIGITAL SIGNATURE ALGORITHM:

A group of Korean cryptographers, in association with government-supported agencies, has been developing a candidate algorithm for Korean digital signature standard, which is named KCDSA (standing for Korean Certificate-based Digital Signature Algorithm). KCDSA is a variant of ElGamal, similar to DSA, and it is designed by incorporating several features from the recent cryptographic research and thus is believed to be secure and robust^[12].

A. Notations Used in The Algorithm

$a \text{ XOR } b$: exclusive-or of two bit strings a and b .
 $allb$: concatenation of two bit strings a and b .
 $Z_n = \{0, 1, n-1\}$; $Z_n^* = \{x | 1 \leq x \leq n-1 \ \& \ \gcd(x, n) = 1\}$
 $|A|$ denotes the bit-length of A for an integer A .
 K belongs to S denote that k is chosen at random over the set S .

B. Signature Generation

The signer can generate a signature $\{r, s\}$ for a message m as follows:

1. Randomly picks an integer k in Z_q^* and computes $w = \text{pow}(g, k) \bmod p$
2. Computes the first part r of the signature as $r = h(w)$,
3. Computes $e = r \text{ XOR } h(z || m) \bmod q$,
4. Computes the second part s of the signature as $s = x(k - e) \bmod q$, and
5. If $s=0$, then repeats the above process

The computation of w is the most time-consuming operation in the signing process. However, since the first two steps can be performed independent of a specific message to be signed, I may pre compute and securely store the pair $\{r, k\}$ for fast on-line signature generation^[16]. The above signing process can be described in brief by the following two equations:

$$r = h[\text{pow}(g, k) \bmod p] \text{ with } k \text{ belong to } Z_q^*$$

$$s = x(k - r \text{ XOR } h(z || m)) \bmod q$$

C. Signature Verification

On receiving $\{m || r || s\}$, the verifier can check the validity of the signature as follows:

1. First checks the validity of the signer's `certi_cate`, extracts the signer's certification data `Cert_Data` from the certificate and computes the hash value $z = h(\text{Cert Data})$.
2. Checks the size of r and s : $0 \leq r < \text{pow}(2, |q|)$, $0 < s < q$;
3. Computes $e = r \text{ XOR } h(z || m) \bmod q$,
4. Computes $w = \text{pow}(y, s) \text{ pow}(g, e) \bmod p$, and
5. Finally checks if $r = h(w)$.
6. Finally checks if $r = h(w0)$.

Limitations

1. It is prone to brute force attack^[14].
2. In this, public key is validated by means of a certificate issued by some trusted third party authority. Therefore, it may be possible that third party may issue wrong public key intentionally or unintentionally.
3. It also requires a collision-resistant hash function which is very difficult and complex to produce. In this, hash code^[15] is not fixed, it is variable

Problem Statement

RSA is a highly secure algorithm. The only known way to attack it is to perform a brute force attacks on the modulus. This attack can be easily defeated to simply increasing the key size [10]. However

A. Increased processing time – as a rough guide, decryption time increases 8-fold as key sizes double.

B. Computational Overheads – the computation required to perform the public key and private key transformations.

Increased key storage requirement – DSA key storage (private keys and public key) requires significant amounts of memory for storage.

Proposed Algorithms

The proposed digital signature algorithm is an adaptation of the RSA Algorithms that overcomes the shortcoming of the RSA Systems (processing time). The new algorithms can solve the problems of processing time while keeping the key size intact, by the making modification of RSA Algorithms.

The new algorithm can solve the problem of processing time by not increase the key size but using key with small bit (1024 bit). So the problem of increase processing time can be solved.

We are trying this algorithm for the implementation of small size data.

There are three important parts:-

1. Key Generation

Suppose a user wishes to send a private message to B over an insecure transmission medium. A & B take the following algorithms to generate a public key and private key. Where P & Q are prime numbers and K is the key.

Input Bit length of modulus, K

Output public key (X, Y), and private key (Z, Y).

1. Generate prime number P_a and P_b of the bit length $K/2$
 2. Generate prime number Q_a & Q_b of bit length $K-(k/2)$
 3. Compute $Y \leftarrow P.Q$ for both a & b
 4. Select X to be an integer, where $(P-1).(Q-1)=1$
 5. Compute Z such that $X.Z = \text{Mod}(P-1).(Q-1)$
 6. Return $(X_a; Y_a)(Z_a; Y_a)$ for A $(X_b; Y_b)(Z_b; Y_b)$ For B
2. *Signature generation*

Input Private Key $(Z_a; Y_a)$ for the sender, public key $(X_b; Y_b)$ for the receiver, the message to be signed, M.

Output s, signature of M

- 1) $A \leftarrow h(M)^{X_b} \text{Mod } Y_b$
- 2) $S \leftarrow A^{Z_a} \text{Mod } Y_a$
- 3) Return(s) A common hash algorithms used in SHA-1

3. Signature Verification

Input Private key $(Z_b; Y_b)$ for the receiver, public key $(X_a; Y_a)$ for the sender, message (M) and Signature (S)

Output VALID or INVALID

1. $B \leftarrow S^{X_a} \text{Mod } Y_a$
2. $Q \leftarrow B^{Z_b} \text{Mod } Y_b$
3. If $Q = h(M)$, Return VALID
Else, Return INVALID

Experimental Results

To compare the performance characteristics of the RSA, DSS and the proposed Algorithms, we developed the program using C language. Then we tested each of the three main components key generation, signature generation and signature verification in each program independently. Tests were performed on an intel P4 3.06 Ghz machine with 512 MB RAM. The Experiments results are tabulated as shown in below table.

Result of Experiment no.1 for text:-

Table: 1

Algo.	Key Gen. (second)	Sign.Gen(Sec ond)	Sign.Ver. (Second)
RSA	4.48490	.017000	13.59500
DSS	34.3270	8.11100	4.532000
Proposed	10.7180	0.01400	0.015000

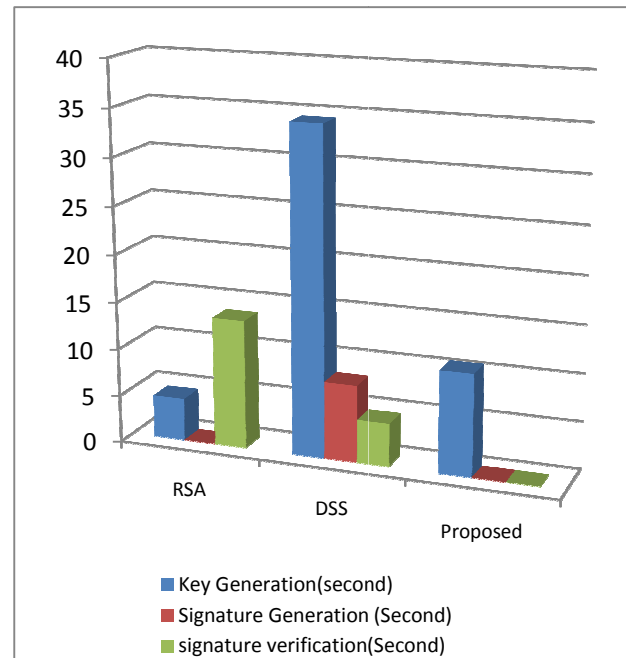


Fig. 2. Result of creating Signature

Result of Experiment no.2 for Audio:-

Table:-2

Algo.	Key Gen. (second)	Sign.Gen. (Second)	Sign.Ver. (Second)
RSA	7.735100	0.16100	13.28110
DSS	72.922100	30.906100	8.313000
Proposed	10.740000	0.030000	0.046000

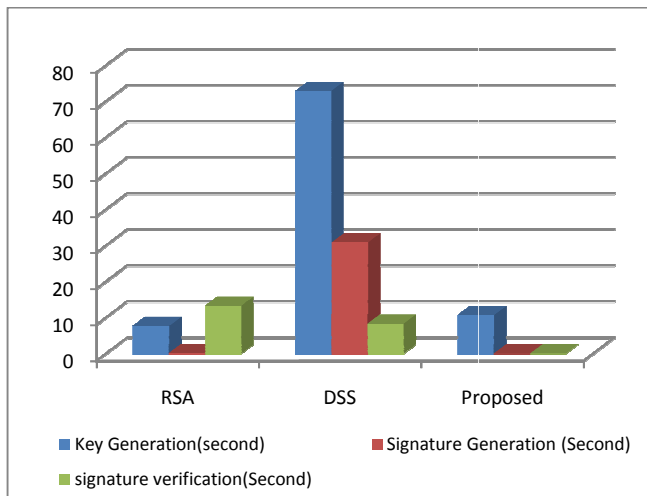


Fig. 3. Result of creating signature.

Result of Experiment no.3 for video:-

Table: 3

Algo.	Key Gen. (second)	Sign.Gen. (Second)	Sign.Ver. (Second)
RSA	12.32500	0.016100	19.62600
DSS	50.64100	35.34410	10.63210
Proposed	11.45400	0.040000	0.050000

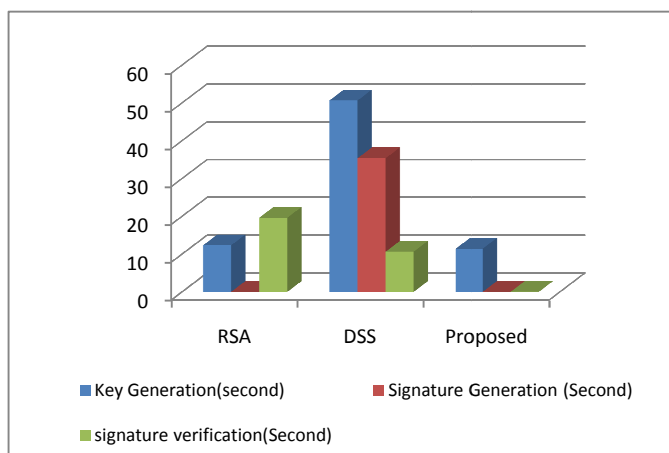


Fig. 4. Result of creating signature

Table: 4

Data Type	RSA	DSS	Proposed
Audio (566 KB)			
Key Gen.	7.735100	0.16100	13.28110
Sig. Gen.	72.922100	30.906100	8.313000
Sig. Veri.	10.740000	0.030000	0.046000
Total	91.3972	31.0971	21.6401
Video (1.38 MB)			
Key Gen.	12.32500	0.016100	19.62600
Sig. Gen.	50.64100	35.34410	10.63210
Sig. Veri.	11.45400	0.040000	0.050000
Total	74.42	35.4002	30.3081
Text (322 KB)			
Key Gen.	4.484900	.017000	13.59500
Sig. Gen.	34.32700	8.111000	4.532000
Sig. Veri.	10.71800	0.014000	0.015000
Total	49.5299	8.142	18.142

4. CONCLUSION

The widespread adoption of Internet as a secure medium for communication and e-commerce has made digital signature implementation to play a vital part of today's information systems. Now-a-days I need I have very large size documents that need to be transferred from one place to another with high security and with minimum time considerations for securing it by the use of digital signature. So, the demand of present scenario is that to develop a secure and efficient implementation of digital signature.

We have developed a digital signature implementation algorithm which exhibits processing power of high performance, efficiency and in minimum amount of time. In this I used private key cryptography for key generation because Public-key cryptography is relatively slow and is only suitable for encrypting small amounts of information while private key cryptography is much faster and is suitable for encrypting large amounts of information. Also I used hexadecimal S-box which implements a much faster processing while encryption and all the work are done in two-dimension rather than one-dimension.

We have applied the most efficient way for providing the highest security but I can further increase by taking larger key size and by using the concept of certificate authority for the distribution of keys. And for the concept of multiuser I can use the concept of different keys at sender's or receiver's end but I have to compromise with the confidentiality.

REFERENCES

- [1] A. Hosseinzadeh Namin, "Elliptic Curve Cryptography, "university of Windsor, April 2005
- [2] Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone, "Handbook of Applied Cryptography", CRC Press, 1997
- [3] Burt Kaliski, "The Mathematics of the RSA Public-Key Cryptosystem", RSA Laboratories. April 9, 2006
- [4] Bruce Eckel, President & MindView Inc, "Thinking in C++", Volume 1&2, 2nd Edition, Completed January 13, 2000
- [5] David Youd, "An introduction to Digital Signatures", published 1996
- [6] [FIPS186] "Digital Signature Standard (DSS)", "Federal Information Processing Standards Publication
- [7] Blake I.F., Van Oorschot P.C. and S. Vanstone. (1986) "Complexity issues for private key cryptography, Performance limits in communication, Theory and Practice", NATO ASI Series E: Applied Science
- [8] Chang C.C., Jan J.K. and Kowng H.C. (1997). "A digital signature scheme based upon the theory of Quadric Residues", *Cryptologia*
- [9] Damgard I.B. (1987). "Collision free hash function and public key signature scheme *Advance in Cryptology*"
- [10] Desmedt Y. (1988). Society and group oriented cryptography, *Advances in Cryptology Crypto*.
- [11] Diffie W. and Hellman M. (1976). New directions in Cryptography, *IEEE Trans. Information Theory*.
- [12] Goldwasser S., Michali S. and Rivest R. (1998). A digital signature secure against adaptive chosen message attacks, *SIAM Journal on Computing*
- [13] Hwang T., Li C. and Lee N. (1993). Remark on the threshold RSA signature scheme, *Advance in Cryptology*
- [14] Merkle R. C. (1987). A digital signature based on conventional encryption function, *Advance in Cryptology*
- [15] D. Johnson, A. Menezes, and S. Vanstone. The elliptic curve digital signature algorithm (ECDSA). *International Journal of Information Security*, 1(1):36–63, August 2001.
- [16] K. Järvinen. Design and implementation of a SHA-1 hash module on FPGAs. Technical report, Helsinki University of Technology, Signal Processing Laboratory, November 2004.

A Semantic Exert to Implement a Novel Framework for E-Learning Web Using Protégé

Akhilesh Dwivedi¹, Aparna Bawankan²

¹Assistant Prof: Department of IT, Dronacharya College of Engineering, Greater Noida, India
dwivedian5@gmail.com

²B.Tech Scholar: Department of IT, Dronacharya College of Engineering, Greater Noida, India
aparnabawankan@gmail.com

Abstract: Semantic E-Learning Web is proving to be a good platform when modern learning & teaching methods come in view. Also Semantic Web can be used to implement e-learning where one can customize learning content as per one's need.

Protégé is one of several open source available tools to implement Ontologies for semantic web. This paper describes an implementation of a novel framework for semantic e learning web along with a significant efforts using Protégé (Protégé 3.5 Beta Version).

Keywords: E-Learning, Semantic Web, Ontologies, Protégé, Semantic E- Learning Web

1. INTRODUCTION

E-Learning expanded as electronic learning is a community which is adapting many modern web technologies like XML,

XMLS (XML Schema), RDF, RDFS (RDF Schema) and many other web technologies from W3C and from other sources at a very rapid rate [1-7]. E-Learning comprises of all e-supported learning and teaching methods [8-9].

Semantic web is an efficient way to implement E-Learning using the components of semantic web like XML, RDF, Ontologies, or OWL (Web Ontology Language) etc. This will make the system both electronically and semantically active and hence will prove to be more beneficial to both instructor and the learner.

2. SEMANTIC WEB

Semantic web is a web service idea provided by Sir Tim Berners Lee. Semantic web means developing an environment in which human and machine agents can work semantically [4, 5, 7]. This web environment can enable machine agents to understand the query asked and revert to the client with the most optimal solution. Although today's web environment gives you the access to surplus amount of information services available on the net, but, it lacks the means to understand the underlying data and hence a

complete, optimal and intelligent solution is not provided to the users [2, 4].

A. Layers of semantic web

XML Layer: XML has its extended form as Extensible Markup Language which is designed to transport and store data. Since XML is not used to display data, hence HTML is embedded with XML for the display of data [6, 7]. XML does not have its tags predefined and the user can define his/her own tags as per the need i.e., the tags in XML can be customized.

Although XML gives the advantage of having user defined tags, it is not enough to make a document semantically active.

RDF Layer: RDF stands for Resource Description Framework which is a language for unfolding information and resources on the web [4, 5, 13]. RDF is a part of W3C's semantic web activity and is a W3C recommendation. RDF is written in XML and is designed to be read and understood by computers [6, 7].

Ontology: Ontology is a Greek word which has its wordily meaning as 'Categories of being and their relations'. Traditionally ontology concerns with existing entities and the relationship among them. We can relate the traditional meaning of ontology with its meaning in Information Sciences as well.

In Computer Science and Information Science, ontology formally represents knowledge as a set of concepts within a domain and the relationship among these concepts [12].

Ontology makes a common platform for any entity which wants to share data or information about a particular domain.

B. Benefits of Developing Ontology [10-13]:

Development of ontologies has its certain benefits in making a system semantically active. Some of them are as follow:

- 1) To share common properties of structure of information among people or software agents.
- 2) To relate between two concepts of same domain.
- 3) To make relationship among various concepts using properties.
- 4) OWL i.e., Web Ontology Language is the language used to represent ontologies.

OWL is the most recent development in standard ontology language from World Wide Web Consortium (W3C) [11].

C. Components of OWL Ontologies:

OWL ontology consists of following components [10-13]:

- 1) Individuals
- 2) Classes
- 3) Properties

1. *Individuals*: Individuals correspond to objects in the domain in which we are interested, also called as domain of discourse. Individuals are also referred as instances or can also be called as being instances of classes [10]. To make two different concepts of same domain relate to each other, OWL does not make use of UNA (Unique Name Assumption), i.e., an individual can have more than one name.

2. *Classes*: In OWL (Web Ontology Language) classes are defined as the type of properties of which the individuals are a part of. OWL uses 'restrictions' to define these classes. A restriction itself is a class like any other class of some individuals [10]. For example: a class of individuals who are instructors, a class of learners who scored Grade A etc.

3) *Properties*: In OWL, properties are used to link together two individuals. To be precise, instances of properties are the one who make relation between two individuals [11]. These properties correspond to associations (relationships). There are mainly three types of OWL properties:

- a) *Object Property*
- b) *Datatype Property*
- c) *Annotation Property*

a) *Object Property*: Object properties are relation between two individuals. It means how two individuals are linked to each other. For example: Adam has wife Eve. Here, in this example, has wife is the property between two individuals-Adam and Eve, defining their relationship. Object property has some following types:

- Inverse Properties
- Functional Properties
- Inverse Functional Properties
- Transitive Properties
- Symmetric Properties
- Asymmetric Properties
- Reflexive Properties
- Irreflexive Properties
- *Inverse Properties*: Inverse property represents the mutual relationship between two individuals i.e., if A is related to B with some relation then B is related to A with the inverse of that relation. For example as shown in Fig. 1, if, 'Adam has wife Eve' is a relation from Adam to Eve then 'Eve has husband Adam' represents the inverse relation from Eve to Adam.

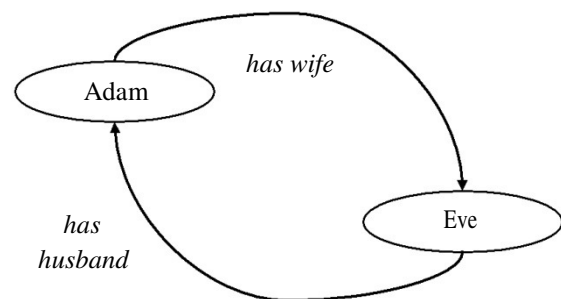


Fig. 1. Inverse Property

- *Functional Properties*: A functional property corresponds to the kind of property where only one individual can be related to the other individual via the property [12]. For example, if we say that C is the mother of A and B then according to Functional Property A and B can have only C as their mother. The property is depicted in Fig. 2. Therefore, for has mother as a relation, an individual can relate to at most one other individual only.
- *Transitive Properties*: Transitive property says that if A is related to B and also, B is related to C, then we can depict that A is related to C as well as shown in Fig. 3.

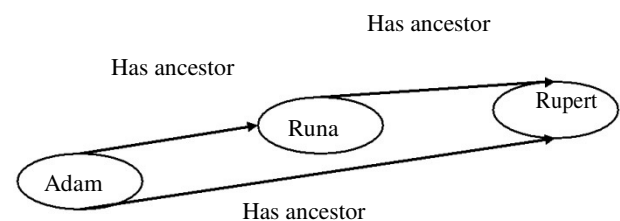


Fig. 3. Transitive Property

- *Symmetric Properties:* As the name suggests, symmetric property is the property where if A is related to B via some property then the inverse property is same as the original property as shown in Fig. 4.

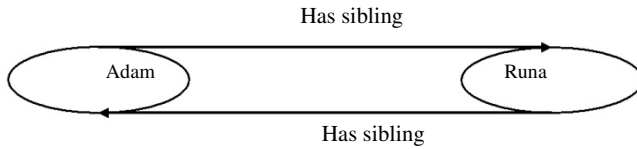


Fig. 4. Symmetric Property

In the above example, if Adam has sibling Runa, then its inverse property must be Runa has sibling Adam.

- *Asymmetric Property:* As the name suggests, Asymmetric Property is the opposite of Symmetric Property where if A is related to B via property P, then B can never have the property P in relation to A as shown in Fig. 5.

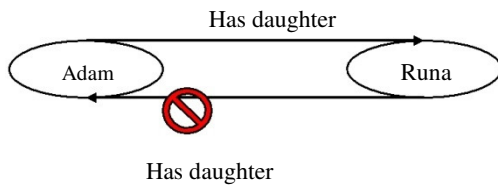


Fig. 5. Asymmetric Property

3. A NOVEL FRAMEWORK FOR SEMANTIC E-LEARNING WEB (THE NEW MODEL)

The new model is shown in Fig. 6 “A Semantic Expert To Implement A Novel Framework For E-Learning Web Using Protégé” is an attempt to implement a semantically active E-

Learning system making use of different semantic web components and e-learning agents. In this framework we are using six E-Learning agents [8, 9]. The names & their functioning are given below in Table I. in brief:

- Instruction Agent
- Lesson Planning Agent
- Resource Location Agent
- Personalization Agent
- Learner Centered Agent
- Collaboration Agent

A. Description of the model

The novel model has made use of e-learning agents however it is a redefined, and much more accurate as compared to the

further implementation of model, description of which is given below:

- 1) The Instructor has the supremacy to authenticate the users i.e., the learners. This gives them the power to add or remove students from the group. On the Learner side any registered student can login to the group or a new user can sign up for the same.
- 2) The Instructor can make tests or quizzes for the learners. This tests and quizzes are received by the student on the learner's side.

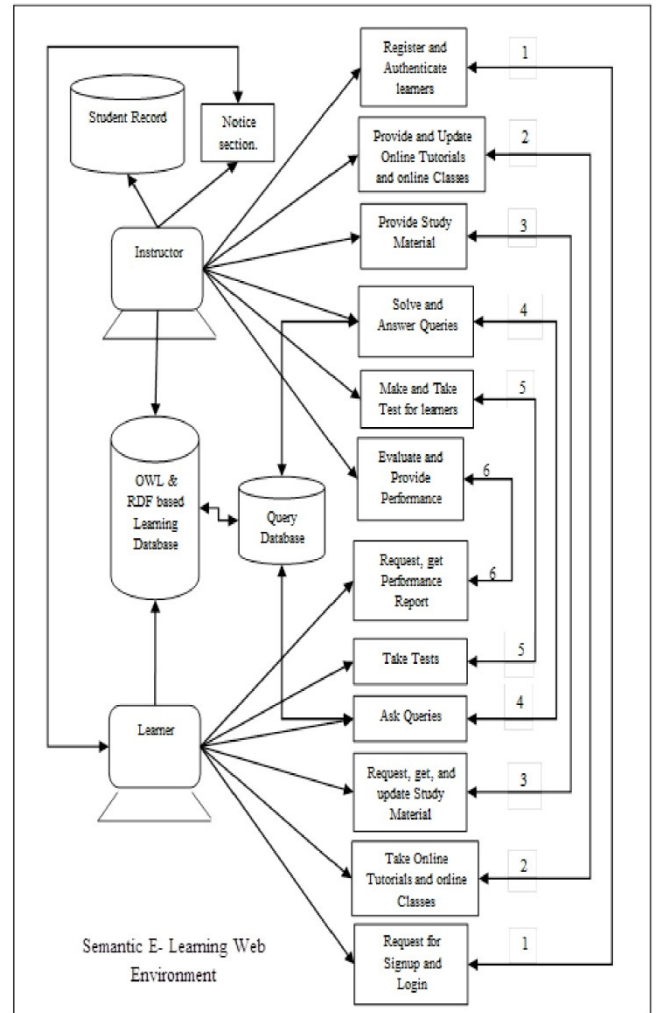


Fig. 6. The new model for A Semantic Expert To Implement A Novel Framework For E-Learning Web Using Protégé

- 3) Based on the tests taken by the learner, the instructor can evaluate learner's performance, the record of which will be provided to the learner. This record can be sent to Learner Centered Agent so that the track of improvement in performance of learner can be kept.

- 4) The instructor can provide learning material to the database. Here Personalization Agent is used to provide the learning material as per learner's prior knowledge and his/her performance. Personalization agent communicates with Learner Centered Agent to do so.
- 5) The learner can ask queries to the instructor through the Collaboration Agent. Collaboration Agent forwards those queries to the instructor and can make available the answer to the learner. These queries are stored in the database along with their answers for future use.
- 6) An instructor can also maintain a database for student record.
- 7) An instructor can make announcements in the notice section regarding placement, timetable, or absentees.
- 8) There is a learning database which is available to both the instructor and the learner. This database can be a library containing videos, PowerPoint slides, e-books etc. Instructor and Learner both can upload data in this database.

Table I : The Names & Their Functioning of Six E-Learning Agents

	Instruction Agent (I.A)	Lesson planning Agent (L.P.A)	Resource Location Agent (R.L.A)	Learner Centered Agent (L.C.A)	Personalization Agent (P.A)	Collaboration Agent (C.A)
Instructor	Alter the learning material Help in deciding among various learning styles	Planning of course Scheduling of learning content	Locate resources on the web Required Learning material to decide the content	Maintain record of learner's performance	Analyze the technique preferred by learner Provide Learning material as per individual's choice.	Get queries of the students Make announcements
Learner			Locate resources on web	Provide feedback Communicate to P.A regarding learner's choice.	Get learning material according to learner's choice by communicating to L.C.A	Forward learner's query to the instructor Get the notice provided by the instructor.

4. PROTÉGÉ

Protégé is an editor for developing ontologies and is freely available [13, 14]. It is being developed at Stanford University in collaboration with the University of Manchester. This application is written in Java and heavily uses Swings to create relatively complex user interface [15]. It provides users with a set of tools to assemble domain models and knowledge based application with ontologies [16].

The Protégé OWL editor enables user to define ontologies which may include classes, their description, properties, and instances [14-17]. There are various versions of Protégé available some of which are- Protégé 4.2 beta, Protégé 4.1 release, Protégé 3.5 beta, Protégé 3.4.8 release and many more.

A. Generating Classes

We make an attempt to develop the ontology for an E-Learning system using only two agents at the beginning- Instruction Agent and the Learner Centered Agent. We make

a class called 'Instruction Agent' and make 'Subject' as its subclass. In order to expand it further we make 'Course' as the subclass of subclass 'Subject'. By doing this we have made a class 'Instruction Agent' which will be serving instructors and will keep a record of the subjects taught by various instructors along with the course material. Fig. 7 shows the completion of above steps. Now we make a sibling class of 'Instruction Agent' called 'Learner Centered Agent'.

Learner Centered Agent is a class keeping record of students i.e., their name, branch, batch, grades etc. Hence we make a subclass of 'Learner Centered Agent' i.e., 'Student Record'. Now in order to maintain the data of all students, we make further subclasses of Learner Centered Agent i.e., Name (storing student name), Branch (stream of study), Batch (year of starting and completion of course), Grades (Grades scored in any particular subject). Since the instructor too should have the data of student, therefore we make 'Student Record' a subclass of Instruction agent as well. Fig. 8 depicts the above steps.

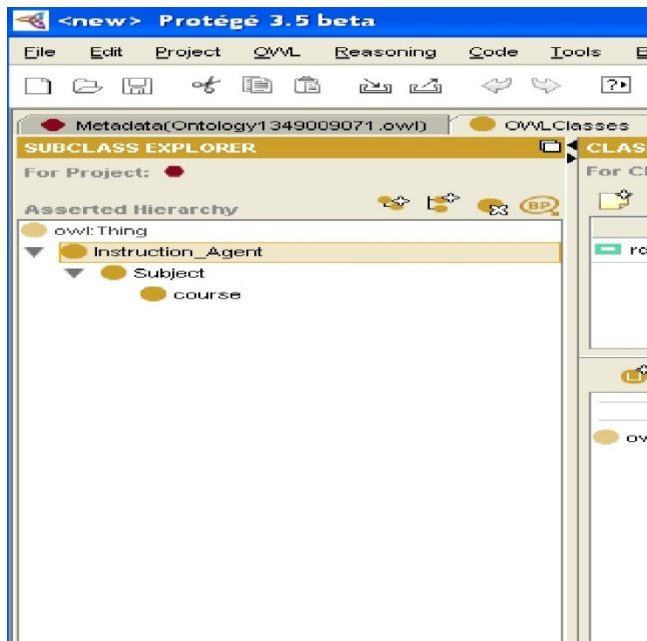


Fig. 7. Generating Instruction Agent in Protégé

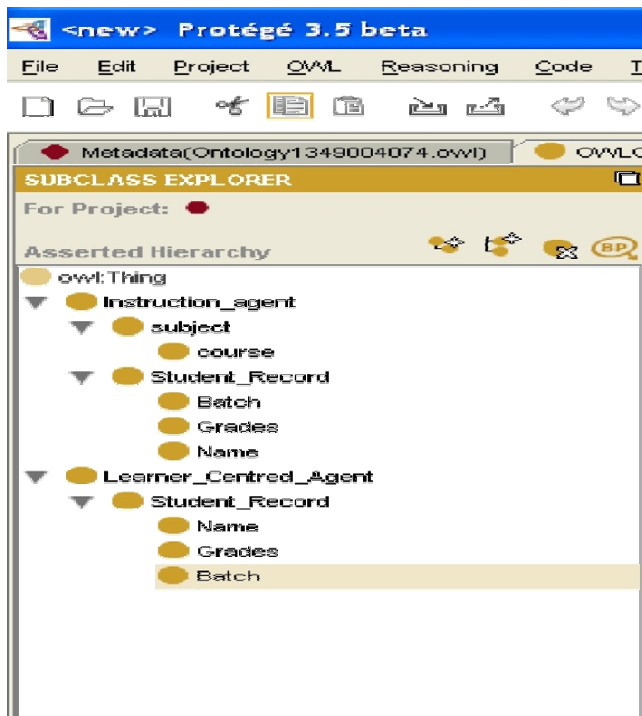


Fig. 8. Class hierarchy of E-Learning system using 'Instruction Agent' & 'Learner Centered Agent'

C. Developing Properties

In Protégé we can develop all the above listed properties of OWL; inverse property is shown below in Fig. 9. and Fig. 10.

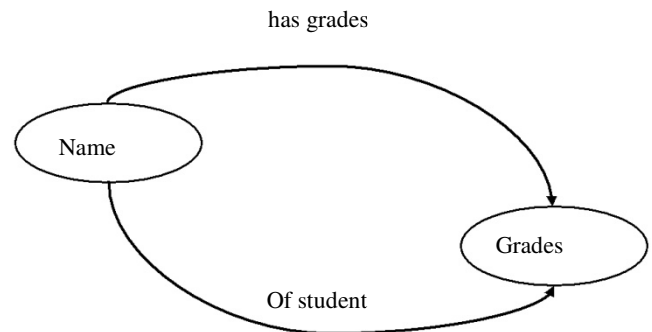


Fig. 9. Inverse Property between student & grades

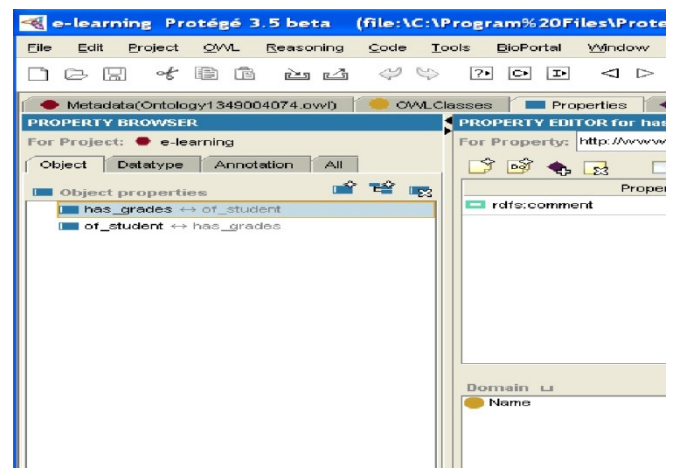


Fig. 10. Inverse Property in Protégé

The above is called an Inverse Property between Name of the student & his/her grades. This means that the name of student and his/her grades are now related with each other. When working on Functional Property, it is a property where only one individual can be related to another via a property. For example: a student id can be associated with only one student. In protégé as well we have the advantage of linking these two via functional property. Functional property in Protégé is shown in Fig. 11. Also transitivity can be shown in Protégé among student name, id, and grades. A student (name of student) will be allotted a unique student id. And grades will be given to that student id as per the performance of the student. And likewise the name of the student is related to the grades via transitive property.

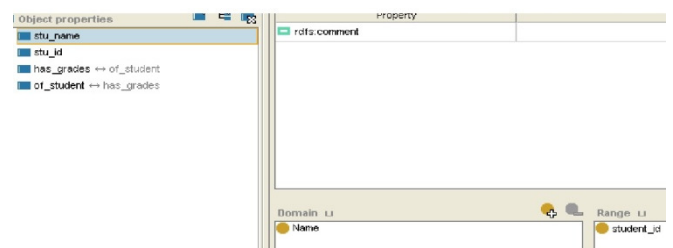


Fig. 11. Functional property in Protégé

5. CONCLUSION & FUTURE WORK

In this paper, we have made an attempt to implement a novel model to make it more accurate and efficient and semantic. However the e-learning agents used are dedicated for the same as allotted to their names in this model i.e., Instruction Agent, Lesson Planning agent, Learner Centered Agent, Personalization Agent, Resource Location Agent and Collaboration Agent. We will try to find some security improvements for valuable data over the Semantic E-Learning Web in near future [19-22]. Above and beyond drawing the new framework, we have also tried to implement the model using Protégé (Protégé 3.5 Beta Version). Protégé is a freely available editor for developing ontologies.

REFERENCES

- [1] Nilsson. M., et al., "Semantic Web Meta-data for e-Learning- Some Architectural Guidelines", Proceedings of the 11th World Wide Web Conference (WWW2002), Hawaii, USA, 2002, pages 1-22.
- [2] Auna Gerber, Alta van der Merwe, Andries Barnard, "A functional semantic web architecture", The Semantic Web: Research and Applications, Lecture Notes in Computer Science Volume 5021, 2008, pp 273-287.
- [3] Lassila, O., van Harmelen, F. ; Horrocks, I. ; Hendler, J. ; McGuinness, D.L. "The Semantic Web and Its Languages," *IEEE Intelligent Systems and their Applications*, Nov. /Dec. 2000, vol. 15, no. 6, Pp. 67-73,
- [4] Guangtao Xue, Qunhua Pan, and MingLu Li, "A New Semanticbased Query Processing Architecture," international Conference on Parallel Processing Workshops, 2007. ICPPW 2007, September 10-14, 2007, Page(s): 63
- [5] en.wikipedia.org/wiki/E-learning. accessed on 27th oct. 2012.
- [6] http://www.w3schools.com/xml/xml_what.asp accessed on 27th oct. 2012.
- [7] http://www.w3schools.com/rdf/rdf_intro.asp accessed on 27th oct. 2012.
- [8] Dawn G. Gregg, "E-learning agents", Learning Organization, Vol. 14 Issue: 4, 2007, pp.300 - 312.
- [9] J. Hendler, "Agents and the Semantic Web," *IEEE Intelligent Systems*, vol. 16, no. 2, Mar. /Apr. 2001, pp. 30-37.
- [10] Knight. C., et al., "An Ontology-Based Framework for Bridging Learning Design and Learning Content", Educational Technology & Society, 9(1), 2006, 23-27.
- [11] [En.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science)) accessed on 24th oct. 2012.
- [12] Noy. F. N., et al., "Ontology Development 101: A guide to creating your first ontology".
- [13] [Owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resource/s/ProtegeOWLTutorialP4_vl_3.pdf](http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resource/s/ProtegeOWLTutorialP4_vl_3.pdf)
- [14] Gennari. J.H., et al., "The Evolution of Protégé: An Environment for Knowledge-Based System Development", International Journal of Human-Computer Studies, Volume 58, Issue 1, January 2003, Pages 89- 123
- [15] Knublauch. H., et al., "The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications" The Semantic Web – ISWC 2004, Lecture Notes in Computer Science Volume 3298, 2004, pp 229-243
- [16] [http://en.wikipedia.org/wiki/Prot%C3%A9g%C3%A9_\(software\)](http://en.wikipedia.org/wiki/Prot%C3%A9g%C3%A9_(software)) accessed on 27th oct. 2012.
- [17] <http://protege.stanford.edu/overview/index.html> accessed on 27th oct. 2012.
- [18] A Dwivedi, S Kumar, A Dwivedi, M Singh, "Current Security Considerations for Issues and Challenges of Trustworthy Semantic Web", International Journal of Advanced Networking Applications (IJANA), Volume: 03, Issue: 01, July-Aug 2011, Pages: 978-983
- [19] D Akhilesh, Kumar S, D Abhishek, S Manjeet, "Cancellable Biometrics for Security and Privacy Enforcement on Semantic Web", International Journal of Computer Applications (IJCA), Vol. no.21, issue no 8, May 2011, pp.1-8
- [20] Akhilesh Dwivedi et al "Defending Against Attacks By Enhancing Security Using Biometrics In Semantic Web", Journal of Global Research in Computer Science, Vol. 2, No 4, 2011, pp.17-28.
- [21] Akhilesh Dwivedi, Abhishek Dwivedi, Suresh Kumar, Satish Kumar Pandey, Priyanka Dabra "A Cryptographic Algorithm Analysis for Security Threat of Semantic E-Commerce Web (SECW) for Electronic Payment Transaction System" Advances in Computing and Information Technology, Advances in Intelligent Systems and Computing Volume 178, 2013, pp 367-379

Comparison between Agile and Traditional Software Development Methodologies & Design A Hybrid Software Development Methodology

Iti Kapoor¹, Prem Sagar Sharma², Atul Kumar Srivartava³

^{1,2}Department of Master of Computer Applications, NIET, Greater Noida

¹itikapoor02@gmail.com, premsagar1987@rediffmail.com

³AITTM, Amity University, Noida – 201303 (INDIA), aksrivastava1@amity.edu

Abstract: During the last several decades, traditional models of software development, called SDLC (Software Development Life Cycle) have been used. These include models like Waterfall, Spiral, Iterative Enhancement, Prototyping etc. However, in the present age of internet and mobile technology, where customer tastes and their requirements change very rapidly, these models have started losing their relevance. In today's era, SDLC is not the appropriate process model to use. Today, as applications move to cloud, and there is fast web-enablement of all services, the development models also has to be in sync with the new realities. Agile SDLC is one such approach. Agile models deliver working software in short time span with short iteration, but almost negligible documentation. However, between these two approaches, there is a need of hybrid agile development model which deals with the shortcomings of both these diametrically opposite models. The proposed hybrid model combines the detailing and depth of a traditional model with the speed of agile model, but with adequate documentation to provide better understanding for developers and users.

Keywords: SDLC, Agile, Iterative, Scrum, Spiral, Hybrid

1. INTRODUCTION

In a rapidly changing world, people's needs are also changing rapidly. The software industry is no different. Every day, different programmers come out in the open presenting what they have done, hoping that the public or their intended group of people will like the software and find it useful. Any software development in itself is a great process that should be followed by any developer for a successful program. Software development life cycle or SDLC is a model of a detailed plan on how to create, develop, implement and eventually fold the software.

A software development process is a structure imposed on the development of a software product. There are several models for such processes, each describing approaches to a variety of tasks or activities that take place during the process. It aims to be the standard that defines all the tasks required for developing and maintaining software. [1]

2. SOFTWARE DEVELOPMENT MODELS

1) Waterfall Model

It is the classic approach and a sequential design process, often used in a software development process. In which, once a phase of development is completed, the development proceeds to the next phase and there is no turning back.

Advantages

1. Fixed requirement.
2. Structured approach, linear so easy to understand and implement

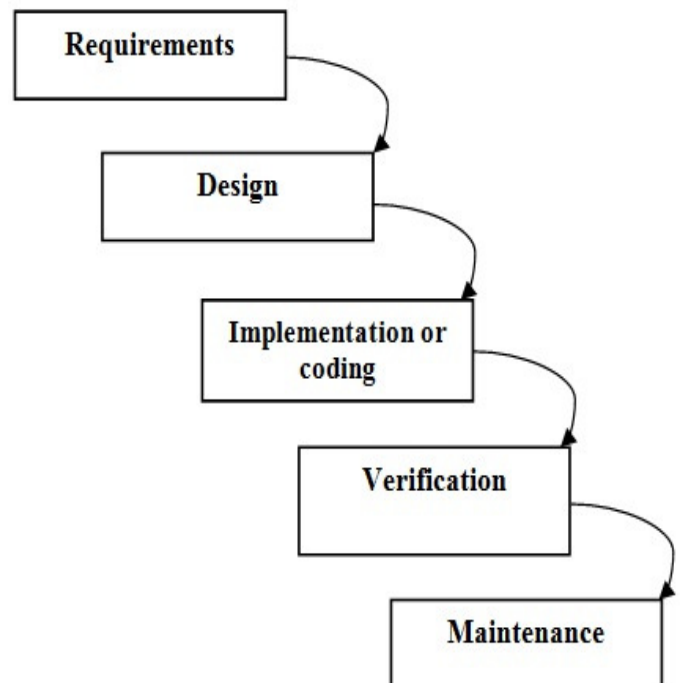


Fig 1. Original Waterfall Model

Disadvantages

1. For long duration projects, requirements may change, acceptability of the product is reduces.
2. Testing is postponed to the later stage till coding is completed
3. Working model is not seen till later stages.

2) Prototype Model

A prototype describe few aspects of, may be completely different from, the final product. It is very useful, when elicitation of requirement is so difficult from customer. It acts as a sample to test the process.

Advantages

1. User involved in the development process and earlier feedback from user.
2. Working model is seen earlier.

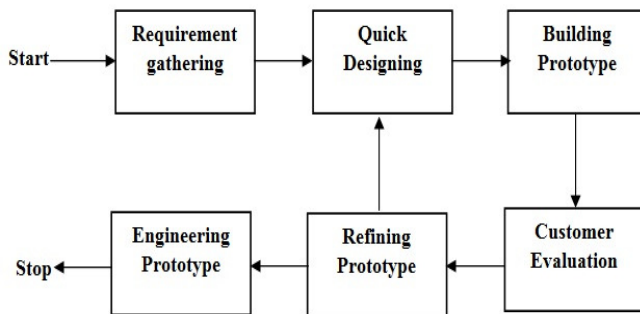


Fig. 2. Prototype Model

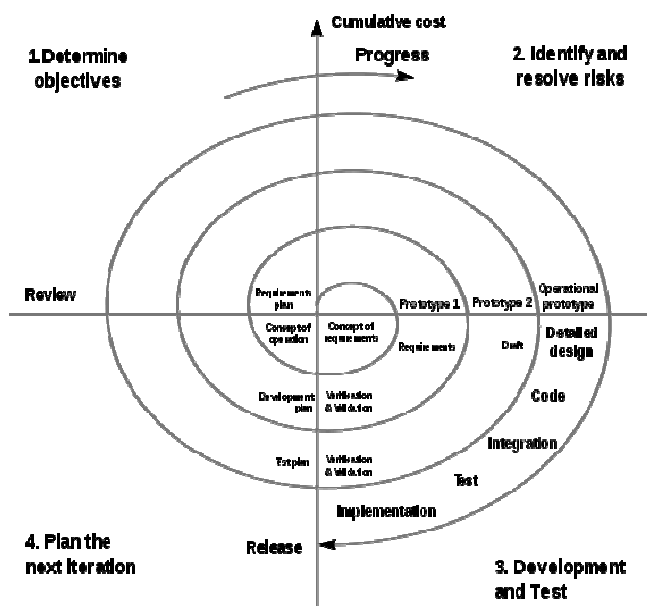


Fig. 3. Spiral Model

Disadvantages

1. Complexity increase
2. Time taken and very costly
3. Spiral Model

This model of development combines the features of the prototype model and waterfall model and focus on risk assessment[1]. This model is intended for large, expensive and complicated projects.

Advantages

1. More reliable
2. Can cope with the changes in requirements.

Disadvantages

Successful implementation depends on the expertise of Risk Management.

4) Iterative Model

In this Model, delivery of software is divided into increments or builds, each increment adding new functionality to the software product and software requirements are well defined[1], but complete product realization could be delayed.

Advantages

Early feedback from customer for better development of product.

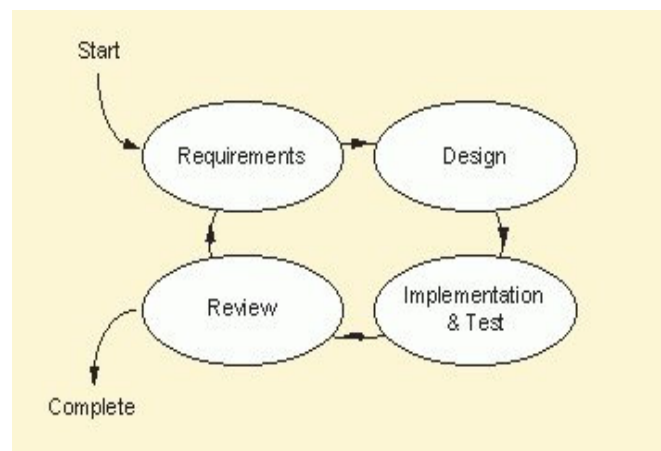


Fig. 4. Iterative Model

Disadvantages

1. More time needed for designing the product.
2. Very costly

Process Approach in Traditional Models

In a typical traditional SDLC model, the following phases and phase-end deliverables are present:

Table1. Process Approach in traditional model

Phase	Deliverables
Start-Up	Project Management Plan (PMP)
Requirements	System Requirements Specification (SRS)
Design	Design Documents
Coding	Unit Tested Code
Testing	Test Plans and Records
Implementation	User Manual, Installation Manual, Release Notes, Training Material

Apart from above, many more documents may be made and some documents may also get combined.

3. AGILE SOFTWARE DEVELOPMENT

Now days, Agile methods are being increasingly used by many IT companies. These methods are gaining in popularity because of low defect rate and higher customer satisfaction. It is mainly used in small software with critical time limit. Agile process requires less planning and it divides the tasks into small increments. Agile process is used for small projects in which changes can be made according to the customer needs leading to satisfaction [4].

Agile methods stress productivity and values over heavy-weight process overhead and artifacts. The Agile Manifesto10, a concise summary of Agile values, was written and signed in 2001 although Agile methods have existed since the early 90s. Agile methods promote an iterative mechanism for producing software, and they further increase the iterative nature of the software lifecycle by tightening design-code-test loop to at least once a day (if not much more frequently) as opposed to once per iteration.

Agile Manifesto (<http://www.agilemanifesto.org>) speaks about following core Values:

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan

Today there are several Agile methods like Scrum, Extreme Programming, Adaptive Software Development (ASD), Dynamic System Development Method (DSDM) etc.

A generalized (ASDM) Agile Software Development Model is given below-

A. Vision and Project Approval

It is a non-iterative phase which is completed in two to three weeks. It mainly deals with the need, scope and boundaries of a new system. Main objective of this phase is to identify critical uses of the system, level of uncertainty of the system, overall estimation of size and duration of the system. Further, systematic analysis is performed to identify the feasibility of the system at operational and economical level with clear specified requirements

B. Exploration Phase

It is an iterative phase, in which meeting are done with the customers to understand the problems with existing system and requirement of new system. New requirements are also defined based on the feedback of last release. Team starts with selection of experienced team members for agile software development. Selected team members start Communicating with the customers to understand the problems and requirements of the proposed system.

C. Iteration Phase

The first activity of this phase is to review the software released in last iteration. Developers and customers do a meeting to prioritize the requirements for the current iteration. In iteration planning, list of requirements in stack is updated depending on the feedback and requirements received from customers.

D. ADCT Phase

This phase deals with the analysis, design, coding and testing. This is the actual implementation phase, which is divided into several iterations. In first iteration, architecture for whole system is developed and then in next iteration designing, coding and testing is performed. Main activities of this phase are simple designing, maintaining coding standards and rigorous testing by Test Driven Development (TDD) and functional testing.

E. Release Phase

This phase is composed of two sub-phases: pre-release and production. In pre-release phase integration and acceptance testing is done and in production phase the developed product is deployed for the customer, training for the user is also done in this phase [5].

Scrum is a process used for agile software development. With Scrum, projects progress via a series of iterations called sprints. Each sprint is typically 2-4 weeks long and sprint planning is essential. While an agile approach can be

used for managing any project, Scrum is ideally suited for projects with rapidly changing requirements such as we find in software development, as shown in figure-

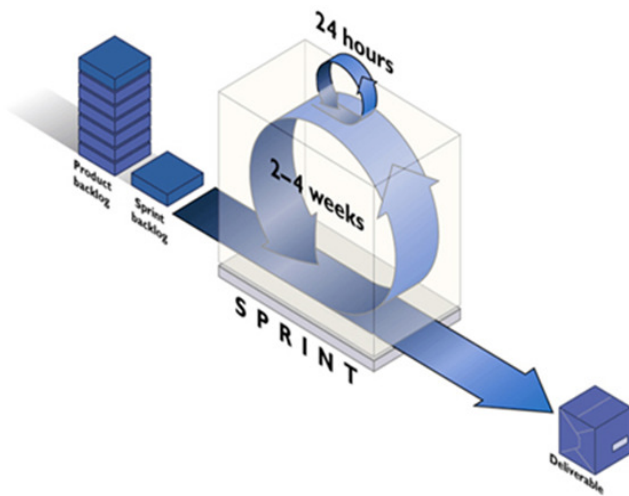


Fig. 5. Agile Model

4. PROCESS APPROACH IN AGILE MODELS:

In a typical Agile model, the following phases and phase-end deliverables are present:

Table 2. Process Approach in Agile model

Phase	Deliverables
Start-Up	Feasibility Study, Product Backlog, Prioritization
Development	Analysis + Design + Coding + testing
Implementation	Pre-release testing + Release for Customer use

Agile project is much better in terms of productivity, performance, faster time cycles, risk analysis. Agile processes are implemented in important applications such as web based, testing tools, etc.

Benefits of the Agile Model

- i. Least documentation
- ii. Faster development
- iii. Product delivered every 2-4 weeks
- iv. Customer contribution in development
- v. Adaptive to the changing environment
- vi. Delivered high quality product

Limitation of the Agile Model

- i. Wastage of resources because of constant changes in requirement
- ii. More helpful for management than developer

- iii. For large projects, effort Estimation is very difficult
- iv. Lack of emphasis on necessary designing and documentation
- v. Experienced people required in development process
- vi. If customer requirement not clear, chance of project failure increases
- vii. CIOs who feel Agile is just another fad or a threat to their cherished hierarchy
- viii. Architects that don't have a clue how to fit applications together in a framework
- ix. Project managers who feel demoted as scrum masters and just won't give up control
- x. Users who are unable to understand their own needs let alone articulate them
- xi. Business system analysts that want to spend 6 months in analysis paralysis
- xii. Developers that are moving slowly and still have software defects
- xiii. Operation management with draconian procedures and inquisitional committees
- xiv. Testing personnel who have no test policy, test plan, agile tools or training
- xv. Help writers that are always the last to know about new features
- xvi. Salesmen that have no idea what the software really does
- xvii. Customer support persons who are totally disengaged
- xviii. Developers that are moving slowly and still have software defects
- xix. Operation management with draconian procedures and inquisitional committees
- xx. Testing personnel who have no test policy, test plan, agile tools or training
- xxi. Help writers that are always the last to know about new features
- xxii. Salesmen that have no idea what the software really does
- xxiii. Customer support persons who are totally disengaged

5. PROPOSED HYBRID SOFTWARE DEVELOPMENT

Which brings us to the fact that neither traditional linear models, nor Agile models are 'fool-proof'. There is no 'best-fit' situation where any of these can be strictly applied. Both these types have their strengths and limitations. What is required is to be able to use the features of both the

approaches, and evolve a new flexi-approach which is practical.

Comparison of Traditional and Agile Models

Table 3. Comparison of Traditional and Agile model

SDLC	Agile	Traditional
Project Plan		✓
SRS (System Requirement Specification)	(Incremental requirement specification)	✓
Design		✓
Coding		✓
User Manual		✓
Test Plan		✓
Test Cases		✓
Installation Manual		✓

Comparing traditional and Agile Models, following hybrid model is proposed.

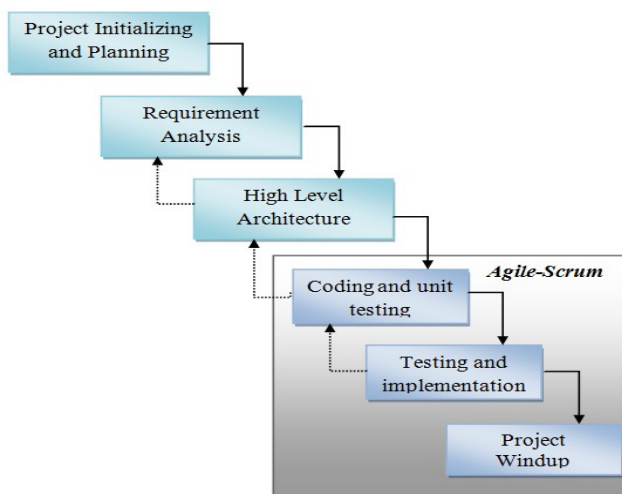


Fig. 6. Proposed. Hybrid Software development Model

As shown in figure, the proposed model is a combination of traditional iterative model and agile scrum.[7] In this model project planning, Requirement analysis, and high level design is done as traditional iterative model and then further processing is done through Agile scrum. This proposed model is useful for large project in distributed environment. This model largely does away with the limitations of Agile, so far as the process is considered, and also brings in the short time-to-market aspect of agile.

In the proposed model, following phases are exist :

Phase 1 (Start-up): This phase includes feasibility analysis and project planning. In this phase the customer involvement is ensured.

Phase 2 (Requirements & Design): The detailed requirements are discussed and agreed. It also needs architecture and high level design of the proposed application.

Phase 3 (Development): This phase is the phase which essentially differentiates between an iterative model and the agile model. In this phase the product features are broken down into smaller components such that each component can be developed using Scrum method. Which means for each component, a cross functional team works with feature backlog, and there is a monthly release to the customer. In iterative model, the independent modules may take a few weeks to a few months to develop. However, in the proposed hybrid model, the modules have to be broken down to such a level, that they do not go beyond 2 – 4 weeks, the planned sprint duration. This is the differentiating feature of the proposed model.

The following table lists the typical phases and deliverable in the proposed model:

Table 4. Process Approach in Hybrid Agile Model

Phase	Deliverables
Start-Up	Project Management Plan (PMP), Feasibility Study
Requirements & Design	Incremental Requirements Specification (IRS), High Level Design
Development	Coding + testing, regression testing
Implementation	Configuration Management + Pre-release testing + Release for Customer use

6. BENEFITS OF THE HYBRID MODEL:

- Monthly release of software to customers
- Development and testing using agile-scrum methodology
- Increased Customer Satisfaction due to faster and better quality delivery
- Flexibility in re-prioritizing the remaining modules
- Necessary documentation for designing
- A broad project plan helping management retain control of the project.
- More documentation that typical agile, helping teams test, install and support and troubleshoot software.

7. CONCLUSIONS

In this paper we have discussed the traditional software development life cycle models, the agile processes, and their advantages & disadvantages. After that, we have compared these two approaches (traditional and agile) and shown the limitations of either approaches. The proposed Hybrid Model[6] is a combination of both traditional and agile scrum and uses the strengths of both. It provides the management control of the project, as well as ensures speed and flexibility from customers' perspective. This model is useful for both small as well as large projects which could be centralised or distributed.

REFERENCES

- [1] Ms. Shikha Maheshwari "A Comparative Analysis of Different types of Models in Software Development Life Cycle" ISSN: 2277 128X Volume 2, Issue 5, May 2012
- [2] Tobin J Lehman, Akhilesh Sharma, "Software Development as a service: Agile Experiences", in annual SRII Global Conference (2011).
- [3] Ahmed, S. Ahmad, Dr. N Ehsan, E. Mirza, S.Z. Sarwar, "Agile Software Development: Impact on Productivity and Quality", in the Proceedings of IEEE ICMIT.(2010).
- [4] Sheetal Sharma, "Agile Processes and Methodologies: A Conceptual Study" IJCSE Vol. 4 No. 05 May 2012.
- [5] S. Bhalerao, D. Puntambekar, "Generalizing Agile Software Development Life Cycle" ISSN: 0975-3397 Vol.1(3), 2009, 222-226.
- [6] <http://www.my-project-management-expert.com/the-advantages-and-disadvantages-of-agile-software-development.html>
- [7] http://blogs.cio.com/michael_hugos/15587/hybrid_agile_moving_agile_development_to_the_next_level

Unique Watermark Generation Using LFSR and EBCDIC Code

Shaikh Rakhshan Anjum¹, Priyanka Verma², Asna Furquan³

^{1,2}Assistant Professor, EXTC, MPSTME, SVKM's NMIMS, Mumbai, India

³Assistant Professor, GGSIPU, Dwarka, New Delhi, India

¹raksha2026@gmail.com, ²priyanka.verma2@gmail.com, ³asn_tech@rediffmail.com

Abstract: In today's multimedia field scenario, the security of watermarked document and the watermark is a crucial issue. The fidelity, integrity and robustness of watermark should be maintained while embedding the watermark into the original document. In this paper we propose a new watermark generation process based on combined LFSR and EBCDIC code technique. In this context, first a pseudorandom bit pattern is generated using LFSR which is not repetitive at all and that pseudo random bit pattern will be encoded by EBCDIC code table. This technique will generate number of unique watermarks. For every pseudorandom sequence we have a different and unique watermark which will be used for watermarking process to secure a document. This technique holds for the copyright protection application, in which if an attacker comes to know about the watermark, he cannot claim for the copyright as he need to decode that watermark to prove his authentication for the copyright document.

IndexTerms: Digital watermarking, LFSR, Seed value, EBCDIC code

1. INTRODUCTION

In the pre-digital era, people's ability to do various things to or with content were limited. The networked digital age makes it possible to do just about anything to digital content, instantaneously and at virtually no cost. While this is a great opportunity for new content business models, it threatens the livelihood of content creators by making rampant piracy possible. Also, more and more public and private entities are going digital and doing business online etc. Thus, we see the need for a technology that enables the secure creation, management, distribution and promotion of digital content on the Internet.

Digital watermarking, also called watermark insertion or watermark embedding represents the scheme that inserts the hidden information into multimedia data, also called the original media or the cover media. The hidden information may be the serial number or the random number sequence, copyright messages, ownership identifiers, control signals, transaction dates, creators of the work, text, bi-level or grey level image, or other digital formats, called the watermark. After inserting or embedding the watermark by specific algorithms, the original media will be slightly modified and

the modified media is called as watermarked media which will be imperceptible from the original media.

The main application for digital watermarking is copyright protection. After embedding the watermark, the watermarked media are sent to the receiver via Internet or other transmission channels. Whenever the copyright of digital media is in question, the embedded information is decoded to identify the copyright owner.

To understand watermarking method and determine their applications, one needs to know the properties of digital watermark. The important properties of watermark are-

(i) **Robustness** of a watermark refers to its ability to survive non-malicious distortions.

(ii) **Data Payload** is the encoded message size of a watermark in a image. The simplest form of watermark has no data payload.

(iii) **Capacity** is the amount of watermark information in the host document. If multiple watermarks are embedded into an image, then the watermarking capacity of image is the sum of all individual watermarks' data payload.

(iv) **Imperceptibility** is the characteristic of hiding a watermark so that it does not degrade the visual quality of image.

(v) **Fidelity** is the visual similarity between the watermarked image and the cover image.

(vi) **Security** of a watermark is the ability of watermark to resist malicious attacks. These attacks include intentional operation of watermark insertion, modification, removal, and estimation which aim at defeating the purpose of watermarks.

(vii) **Computational Cost** is the measure of computing resources required to perform watermark embedding and detection processes.

In this paper, the aim is to introduce a watermark generator which will produce unique watermark depending upon the LFSR pseudo random number generator and the EBCDIC code.

2. LINEAR FEEDBACK SHIFT REGISTER

LFSR is a pseudo random number generator. An LFSR is a shift register that, when clocked, advances the signal through the register from one bit to the next most-significant bit (see Figure 1). Some of the outputs are combined in exclusive-OR configuration to form a feedback mechanism. A linear feedback shift register can be formed by performing exclusive-OR on the outputs of two or more of the flip-flops together and feeding those outputs back into the input of one of the flip-flops[8]. Linear feedback shift registers make extremely good pseudorandom pattern generators [2]. When the outputs of the flip-flops are loaded with a seed value (anything except all 0s, which would cause the LFSR to produce all 0 patterns) and when the LFSR is clocked, it will generate a pseudorandom pattern of 1s and 0s. A maximal-length LFSR produces the maximum number of PRPG patterns possible and has a pattern count equal to $2^n - 1$, where n is the number of register elements in the LFSR. It produces patterns that have an approximately equal number of 1s and 0s and have an equal number of runs of 1s and 0s.

3. 8- BIT LFSR RNG

In this section we discuss about the 8 bit LFSR and its pseudo random generation technique. The most common way to implement a random number generator is LFSR. Codes generated by the LFSR are actually pseudo random sequences because the sequence repeats itself after a certain number of cycles. It is known as the period of the generator. LFSR is based on the recurrence equation,

$$x_n = x_{n-1} \oplus x_{n-2} \oplus \dots \oplus x_{n-m} \dots \text{eq.} \quad (1)$$

The operator \oplus is the exclusive OR (XOR) operator. The equation (1) shows that n th bit can be generated utilizing m previous values with XOR operators [2]. The value of m determines the period of the generator. The achievable maximum period is $2^m - 1$. For the 8-bit LFSR, the recurrence equation is,

$$x_n = x_{n-2} \oplus x_{n-3} \oplus \dots \oplus x_{n-8} \dots \text{eq.} \quad (2.)$$

Since new value x_n depends on previous m values, it is necessary to store previous m values to find the new value. This can be done with m single bit shift registers comprised with flip flops. According to the equation (2), XOR feedback tap positions are taken at 0th, 4th, 5th and 6th flip-flops. The maximum period of the generator is $2^8 - 1$ (255). In each clock pulse, generated new bit is inserted to shift register while the oldest bit shifts out. Output of the 8 flip-flops form

the 8-bit random number. A group of flip-flops connected in series are used with XOR gates to construct the LFSR random number generator as shown in Fig.1. The 8-bit LFSR will generate random numbers which will be coded to give a secret code word. If these random numbers won't match at the extraction process, the original message won't be retrieved.[7]

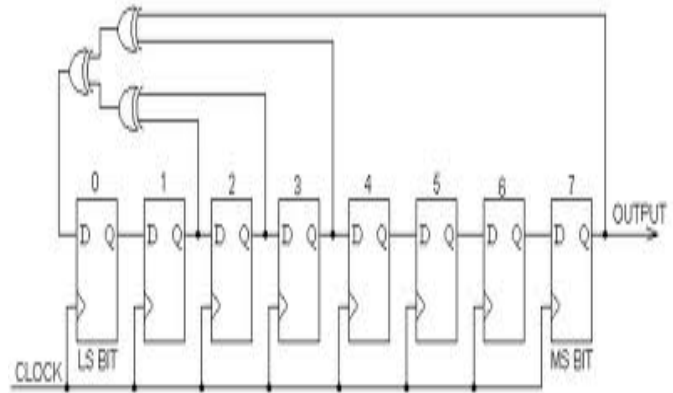


Fig. 1 8 Bit LFSR random number generator

4. EBCDIC CODING

EBCDIC is also known as Extended BCD Interchange Code. It is the code for text files that is used in IBM's OS/390 operating system and that thousands of corporations use for their legacy applications and databases. In an EBCDIC file, each alphabetic or numeric character is represented with an 8-bit binary number (a string of eight 0's or 1's). 256 possible characters (letters of the alphabet, numerals, and special characters) are defined. It is now used for security purposes mainly in the applications of cryptography. EBCDIC code table is shown in Fig.2.

		Low Order Bits															
		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
High Order Bits	0100	space										↓	.	<	(+	
	0101	&										!	\$	*)	:	~
	0110	-	/									;	,	%	-	>	?
	0111											'	"	#	@	'	"
	1000		a	b	c	d	e	f	g	h	i						
	1001		j	k	l	m	n	o	p	q	r						
	1010		~	s	t	u	v	w	x	y	z						
	1011																
	1100	{	A	B	C	D	E	F	G	H	I						
	1101	}	J	K	L	M	N	O	P	Q	R						
	1110	!	S	T	U	V	W	X	Y	Z							
	1111	0	1	2	3	4	5	6	7	8	9						

Fig. 2. EBCDIC code table for encoding

For eg if we want to decode 'SAFE' using EBCDIC coding, then it will be decoded as '11100010110000011100011011000101'.

5. ALGORITHM FOR GENERATING WATERMARK USING 8 BIT LFSR

The Algorithm for generating watermark is shown in Fig.3. Initially a seed value is defined before implementing the algorithm. By implementing this algorithm, total number of 256 pseudo random bit patterns of 8 bit each will be generated.

```

Initialize s state of LFSR.
Initialize t Tap positions.
Define n=length(s) and m=length (t).
C=[states of LFSR] and seq = generated sequence
for k=1:2^n-2
    b(1) = xor(s(t(1)), s(t(2)))
    if m>2;
    for i=1:m-2;
        b(i+1)=xor(s(t(i+2)), b(i));
    end
    end
    j=1:n-1;
    s(n+1-j)=s(n-j);
    s(1)=b(m-1);
    c(k+1,:)=s;
end
seq=c(:,n)

```

Fig. 3. Algorithm for generating watermark using LFSR

6. RESULTS

The 8 bit pseudo random sequence is generated by implementing above algorithm in MATLAB v2012. The generated sequence is shown in fig1.

After the generation of random sequence, the 8 bit pseudo random sequence is concatenated together and encoded using EBCDIC code. The EBCDIC coded sequence is then converted into an image of .bmp format. This image is then used as a watermark for copyright protection. A unique watermark can be created by using different seed values.

```

11100010111001011101001011010100011110110100010010000001101010111010100110010
011101010011100010

```

Fig. 4. Generated sequence using LFSR random number generator.

SVKM'S NMIMS

Fig. 5. EBCDIC encoded watermark

7. CONCLUSION

From the above implementations, it has been concluded that even if any attacker extracts the watermark, he/she cannot decode it and also since the watermark is coded using EBCDIC code the watermark becomes more robust against channel noise. This gives double security to the watermark and hence can be widely used in copyright protection applications.

8. FUTURE SCOPE

In the future work, this technique can be used in applications other than copyright protection like multiple watermarks can be generated and can be embedded into a single image and further transmitted. When extracted by an attacker he/she gets multiple watermarks from which he/she can't make out the original one since the knowledge of the length of the watermark is known only to the desired receiver. Only from the knowledge of the length of the watermark, the receiver extracts the original information.

REFERENCES

- [1] H.C Huang, H.M Hamg and J.S Pan, "Intelligent Watermarking Techniques", ISBN 981-238-955-5.
- [2] W.A.S Wijesinghe, M.K Jacaranda and D.U.J Sonnadara, "Hardware Implementation of Random Number Generators," proceedings of the Technical sessions, 22 (2006) 28-38, Institute of Physics-Sri Lanka, pp.28-38
- [3] Yanqun Zhang, "Digital Watermarking Technology: A Review," IEEE International Conference on Future Computer and Communication, 2009.
- [4] L.Robert, T.Shanmugapriya, "A Study on Digital Watermarking Techniques", International Journal of Recent Trends in Engineering, Vol.1, No.2, May 2009.
- [5] I J Cox, M.L.Miller, A Mckellips. *Watermarking as Communications with Side Information*. Proceeding of IEEE, 1999, 87(7):1127-1141.
- [6] I.J.Cox, M.L.Miller, J.A.Bloom, J.Fridrich, T.Kalker. *Digital Watermarking and Steganography*, Morgan Kaufmann publishers, Seattle, Washington, USA, 2008.
- [7] Xiangxue Li, Dong Zheng, and Kefei Chen, "LFSR-Based Signatures with Message Recovery," International Journal of Network Security, Vol.4, No.3, pp.266-270, May 2007
- [8] John Koeter, "What's an LFSR?" Texas Instruments, SCTA036A, December 1996

Enterprise Portal with Java Portlet for Universities

Veenita Gupta¹, Neeraj Kumar², Seema Rawat³, Praveen Kumar⁴

^{1,2}M.Tech (CS&E) - 2nd Year, ^{3,4}Assistant Professor
¹veenitagupta2009@gmail.com, ²neerajkumar1989@gmail.com
³pkumar3@amity.edu, ⁴srawat1@amity.edu
^{1,2,3,4}Amity University Noida

Abstract: Today many universities focus on the hardware development, but ignore to develop software. The development of the universities will be there if universities will focus on software development also. In this we will deploy enterprise grade university portal with portlet extensibility and enhance the communication and automation in university. In this paper we will see how the portlet will be helpful to the universities in automation process, communication and many other important works of the universities. The requirements of such a portlet is, make use tool for developing the portlet that is uportal and the platform over which we will develop the portlet and its supporting components.

Keywords: Portal; University Portal; uPortal; portlets; portal technology;

1. INTRODUCTION

Many universities in education development are still in the situation of low-level development and redundant investment. The development of the universities will be there if universities will focus on software development also. In this we will deploy enterprise grade university portal with portlet extensibility and enhance the communication and automation in university. Portals serve as a starting point to information and applications on the Internet or from an intranet. Portals provide aggregation of content from diverse sources. To accommodate the aggregation and display of such diverse content, a portal server must provide a framework that breaks the different portal components into portlets. Each portlet is responsible for accessing content from its source and transforming the content so that it can be rendered to the client. A portlet is responsible for providing application logic or storing information associated with a particular user.

In this paper we will see how the portlet will be helpful to the universities in automation process, communication and many other important works of the universities. The requirements of such a portlet is, make use tool for developing the portlet that is uportal and the platform over which we will develop the portlet and its supporting components. uportal is free and open source java implemented web portal platform developed and maintained. portal can aggregate and present content generically for unauthenticated users. When users log in, uportal can do

much more, with personalised content, customized layouts, group membership and access control.

Uportal contains many components in it which are JDK (Java Development kit), ANT, Maven, *HSQldb* and tomcat. These are important components of uportal with out these components uportal can not work. JDK is necessary because a Java Virtual Machine (JVM) is not sufficient for building and running uPortal, JDK includes important platform tools like the compiler. The uportal provides several setup and administrative functions through Ant. Building the uPortal software has been largely migrated to Maven. For the present, these 2 tools work together to handle build, deployment, and some management features of the portal. UPortal, we must have web container. A web container is a JEE server component that enables access to servlets and JSP pages. Tomcat is the easiest implementation to use with uPortal and is recommended for use with uPortal. Portlet development is done using uportal and the various functionalities performed by portlet is described here and how the portlet will be helpful to the universities is can also be seen.

2. EXISTING PROBLEM ANALYSIS

Universities in education development are still in the situation of low-level development and redundant investment. Universities focus on the hardware development, but ignore to develop software. The development of the universities will be there if universities will focus on software development also. Some problems are :-

- *Data Inconsistency*

Data from different systems can not be shared and exchanged, so the data are inconsistent in different department and so on. Inconsistency may lead to lose to the university at large scale if not taken care soon.

- *Lack of Unified Planning*

Various departments build own applications according to their own ideas, and lack effective integration. As a university each department should be properly integrated with all other departments.

- *Time consuming*

Due to not having any common way to data sharing and communication it may lead to wastage of time also.

3. FEATURES OF UNIVERSITY PORTAL

Customization: Every user has complete control over the information presented within the portal. Users can control over what information is and is not displayed on their portal pages.

Single sign-on (SSO): University Portal is a one-stop client-oriented web site. Only one account, the visitor can get everything they need. There will not be data inconsistency as single account will be there.

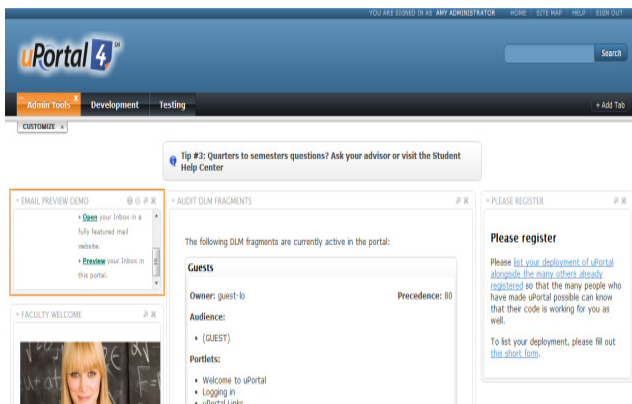
Resource highly integrated: All resources were integrated into one website so that we need not to create many websites. There will be unified planning as integrated resources are available.

Time saving: - *communication* will be easy and data can be shared easily when communicating through single channel.

4. DEVELOPMENT OF PORTLET

For developing portlets following requirements needs to be fulfilled. For this we need an free and open source java implemented web portal platform that is uportal and the platform over which the portlet will be designed along with its supporting components.

A. uPortal



uPortal is an open-standard portal being developed by institutions of higher education of America. uPortal is the leading open source enterprise portal framework built by and for the higher education community. uPortal continues to evolve through contributions from its global community and is supported by resources, grants, donations, and memberships fees from academic institutions, commercial

affiliates, and nonprofit foundations. uPortal is built on open standards-based technologies such as Java and XML, and enables easy, standards-based integration with authentication and security infrastructures, single sign-on secure access, campus applications, web-based content, and end user customization. It is one of the most widely deployed open source enterprise portal frameworks, having been adopted by hundreds of institutions and the Research community, world wide.

B. NetBeans

NetBeans IDE is a free, open-source, cross-platform IDE with built-in-support for Java Programming Language. NetBeans IDE is an open-source integrated development environment. NetBeans IDE supports development of all Java application type. Among other features are an Ant-based project system, Maven support.

The platform offers reusable services common to desktop applications, allowing developers to focus on the logic specific to their application. Among the features of the platform are:

- User interface management (e.g. menus and toolbars)
- User settings management
- Storage management (saving and loading any kind of data)
- Window management
- Wizard framework (supports step-by-step dialogs)
- NetBeans Visual Library
- Integrated development tools

Requirements of Netbeans

- NetBeans Portlet Plugin on NetBeans IDE 5.5
- Project Glassfish
- OpenPortal Portlet Container

The Portal Pack plugins support full life-cycle of portlet application development inside NetBeans. Using this tool portlet developers can develop, package, deploy, and test portlets inside their NetBeans IDE. Portal Pack container has a set of plugins for Netbeans IDE which provide portlet development and deployment support inside Netbeans IDE. In this we need to configure the portlet container with the netbeans.

Purpose of plugin

- You can create JSF based portlet in NetBeans 6.1 IDE using Portal Pack 2.0.

- Portal Pack plugins provide a tight integration with portal servers such as, the OpenPortal Portlet Container on java.net to support deployment and undeployment of portlets on both the local and remote servers.

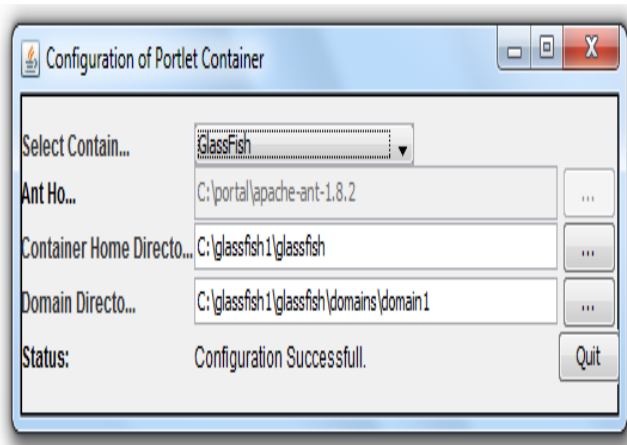


Fig Configuring Portlet container in glassfish

Portlet

Introduction

Portlets are modular panes within a web browser that surface applications, information, and business processes. Portlets can contain anything from static HTML content to Java controls to complex web services and process-heavy applications. Portlets can communicate with each other and take part in Java page flows that use events to determine a user's path through an application. A portlet is a window in the portal that provides a specific service or information from an application development perspective; portlets are pluggable modules that are designed to run inside a portlet container of a portal server. Portlets rely on the portal infrastructure to access user profile information, participate in window and action events, communicate with other portlets, access remote content, lookup credentials, and to store persistent data. Portlet applications provide the means to package a group of related portlets that share the same context.

Portal Technology

Portal is a Web-based integrated information system which takes "application integration" and "Elimination of information silos" as the ultimate goal, provides many functions such as single sign-on, content aggregation, personalization features of the portal. Based on the performance level of a Web application framework as the principle, Portal fully supports the industry standard portal platform framework for Portal Framework, can connect and integrate with different applications or data sources, then show the personalized integration results to the end-user, can dynamically deploy the follow-up service application to

application service platform, and has ability to read, write, update and control the components, handles requests from clients, calls their personalized page for each user. Portlet container provides the Portlets runtime environment, call Portlets through the Portlet API, Portlet container can be called from Portal through Portlet InvokerAPI, which returns the Portal-related information in the Portlet Provider SPI (Service Provider Interface).

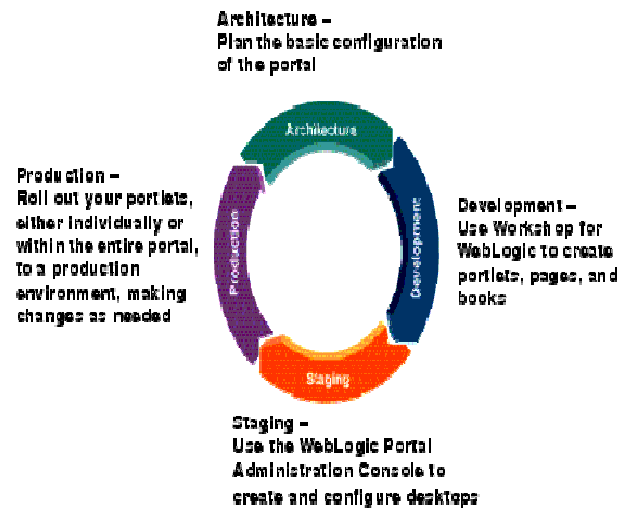


Fig. Portlets and the Four Phases of the Portal Life Cycle

Portlet Description

A portlet is a window in the portal that provides a specific service or information from an application development perspective, portlets are pluggable modules that are designed to run inside a portlet container of a portal server. In this project a uPortal 4 has been used to deploy portlets and further ease the manual work with in universities. The portlets are pluggable modules and are deployed according to adapter JSR 168 and adapter JSR 286. We have developed portlets named as "Snooper Portlet" and "Toggle Resource Aggregation".

Snooper Portlet

Snooper portlet was basically designed to view the request for attributes and headers. Snooper Portlet configuration mainly includes Http Request, Http Header Info and Locales.

The parameters required for Http Request are request protocol, request method, request URI, path translated, query string etc. Http header includes connection, host, accept language, accept, cookie, user agent, referrer etc. Locales includes various languages with their country name. The snooping portlet allows the request attributes and header to

be viewed and it contains a portlet timeout of type portlet. In the configuration of snoopers there is access permission for editing, help and about information. Snooping portlet can be used in development and portlet administration. To view the http request snoopers portlet carries information about request protocol, request method such as GET. The request URI along with server name and server port (8080) are used for rendering the translated path to the requested attribute or header.

Toggle Resource Aggregation

The toggle resource aggregation portlet is used to provide a means to enable/disable CSS and javascript aggregation. The parameter within display settings of toggler are display icon URL, portlet window chrome, theme.

Toggle resource aggregation is used for summary information of uPortal along with toggle CSS and javascript aggregation portlet during registering of new portlet. After registering the new portlet we need to define the categories to which toggler belongs for example development. We can add people and groups to the development category for example portlet administration. Toggler helps in providing lifecycle management/automatic publishing/automatic expiration of new portlet.

Registering Portlet to uPortal

After creating portlet there is need to register the portlet to uPortal. Portlets are registered to uPortal and various properties are also set for the portlet in uPortal. During registration firstly we have to define the type i.e., whether it is CMS, web proxy portlet, bookmark portlet, portlet etc. Here we define the type as portlet to register the toggle portlet.

We have to define the portal in which the portlet is to be kept and then need to give its description in brief, like here we have given the category as uPortal and in description it is given as toggle css/resource aggregation.

We need to define various fields of portlet which we want in it like portlet title, portlet name, portlet description etc. We need to define in which category portlet will lie like development, research, finances, services etc. Then define the people or group of people who can add portlet to their layout for example student are allowed or only for faculty or for portlet administrator only. Then there is a need to define the lifecycle of the portlet like created, published or approved etc.

We need to define the publishing and expiration dates of the portlet along with the time of publishing and the time of expiration. Then portlet with the defined configuration will be displayed.

Snooper Portlet

Configuration of Snooper Portlet



PORTLET ADMINISTRATION

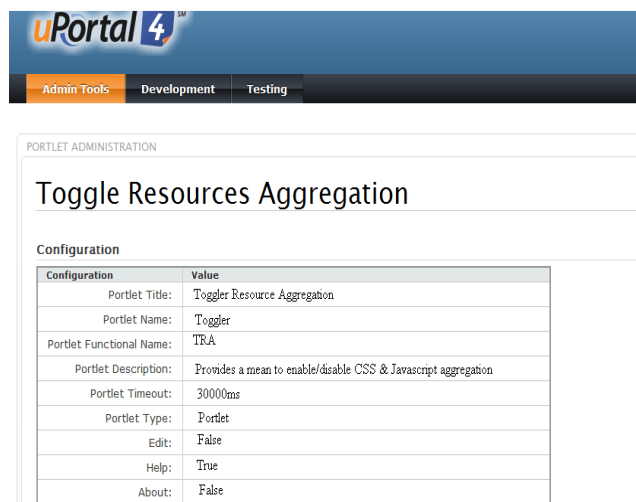
Snooper

Configuration

Configuration	Value
Portlet Title:	Snooper
Portlet Name:	Snooping Portlet
Portlet Functional Name:	Attribute and header request viewer
Portlet Description:	Tools for viewing request attributes and header
Portlet Timeout:	5000ms
Portlet Type:	Portlet
Edit:	False
Help:	True
About:	False

Toggle Portlet

Configuration of Toggle Resource Aggregation



PORTLET ADMINISTRATION

Toggle Resources Aggregation

Configuration

Configuration	Value
Portlet Title:	Toggler Resource Aggregation
Portlet Name:	Toggler
Portlet Functional Name:	TRA
Portlet Description:	Provides a mean to enable/disable CSS & Javascript aggregation
Portlet Timeout:	30000ms
Portlet Type:	Portlet
Edit:	False
Help:	True
About:	False

Registration of Portlet

Register New Portlet

Summary Information

Option	Setting
Portlet Title:	ToggleResourcesAggregation
Portlet Name:	Toggler
Portlet Functional Name:	TRA
Portlet Description:	Provide means to enable/disab
Portlet Timeout:	30000 ms

Controls

Portlet Controls

☒ hasHelp

☐ editable

☐ hasAbout

[Back](#) [Next](#) [Cancel](#)

Setting details of Portlet

Register New Portlet

Lifecycle Management

Option	State	Description
<input checked="" type="radio"/>	Created	Created
<input type="radio"/>	Approved	Reviewed and approved for production
<input type="radio"/>	Published	In production
<input type="radio"/>	Expired	Might need it later, but not using it now

Automatic Publishing (optional)

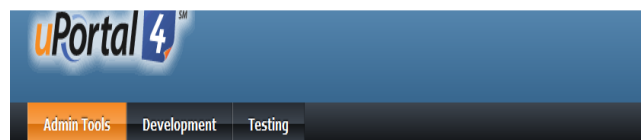
Option	Setting
Automatic Publish Date and Time	07/02/2012 2:00 PM (Reset)

Automatic Expiration (optional)

Option	Setting
Automatic Expiration Date and Time	06/02/2013 2:00 PM (Reset)

Back Next Cancel

Portlet After Registration



PORTLET ADMINISTRATION

Toggle Resources Aggregation

Configuration

Configuration	Value
Portlet Title:	Toggle Resource Aggregation
Portlet Name:	Toggle
Portlet Functional Name:	TRA
Portlet Description:	Provides a mean to enable/disable CSS & Javascript aggregation
Portlet Timeout:	30000ms
Portlet Type:	Portlet
Edit:	False
Help:	True
About:	False

5. CONCLUSION

uPortal is built on open standards-based technologies such as Java and XML, and enables easy, standards-based integration with authentication and security infrastructures, single sign-on secure access, campus applications, web-based content, and end user customization. The campus information portal integrates and manages the existing information systems and Web applications, providing unified access to the entrance, thus change the access patterns of existing information systems. At the same time, bypassing the inherited custom content, applications and services through the customized portal channel, system allows remote users, students, faculty and staff access to campus information portal from the outside, and no additional client software installation and maintenance, reduces portal implementation and management costs, improves work efficiency, will have a major impact on the campus information technology.

REFERENCES

- [1] File:///C:/Users/Dell/Desktop/summer/training/uportal/Build,%20Install,%20&%20Run%20uPortal%204.0.0%20%20%20Unicon.htm
- [2] http://contrib.netbeans.org/portalpack/UG_PortletContainerPlugin.html#mozTocId306290
- [3] http://java.sun.com/developer/technicalArticles/J2EE/sdk_nbp/ortletplugin/
- [4] <http://netbeans.org/kb/articles/portalpack.html#ghlmp>
- [5] <http://netbeans.org/kb/docs/java/quickstart.html>
- [6] <http://en.wikipedia.org/wiki/NetBeans>
- [7] <http://www.wsu.edu/portal-project/background.htmlhttps://www.google.co.in/search?sugexp=chrome,mod=4&sourceid=chrome&ie=UTF8&q=ide+used+for+portlet+development>
- [8] http://contrib.netbeans.org/portalpack/UG_PortletContainerPlugin.html#moz.
- [9] Portal specification JSR 168 and API, <http://www.jcp.org/en/jsr/detail?id=168>.
- [10] WSRP Specification 1.0 by OASIS, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrp.
- [11] A. Akram, D. Chohan, X.D. Wang, D. Meredith and R. Allan, CCLRC Portal Infrastructure to Support Research Facilities, *GGF14*, Chicago, USA, 2005..
- [12] R. Crouchley, A.Fish, R.J. Allan and D. Chohan *Sakai Evaluation Exercise*. Reportto JISC (December2004), http://www.grids.ac.uk/Sakai/sakai_doc.pdf.

Attacks and its Security Mechanism in AODV for Mobile Ad hoc Networks

Mahak Gupta¹, Kushagra Agrawal², Rajneesh Kumar Gujral³, Sanjeev Rana⁴

^{1, 2, 3} Department of Computer Science & Engineering

⁴ Department of Information Technology

M.M. Engineering College, M.M. University, Mullana, Ambala (133207)

¹Emailme.mahak@gmail.com, ²agrawal_kushagra@rediffmail.com,

³dr Rajneeshgujral@yahoo.com, ⁴dr.sanjeevrana@yahoo.com

Abstract: Mobile Ad Hoc Network (MANET) is a multi-hop wireless network of mobile nodes, which forms a temporary network without any help from established infrastructure or centralized administration. Routing in MANET is a challenging task which receives great amount of attention from researchers. For most existing routing protocols of mobile ad hoc network (MANET), more efficient security mechanisms against the attacks from malicious, compromised and selfish nodes are highly demanded. ADHOC On-demand Distance Vector (AODV) is one of the most famous routing protocols in MANETs. In this paper we present a survey for some existing attacks and their countermeasure in AODV for mobile ad hoc network.

Keywords: MANETs, AODV, attacks, security.

6. INTRODUCTION

An ADHOC network is a collection of mobile nodes and to connect these wireless nodes there is a wireless communication network[1]. This type of network is known as MOBILE ADHOC NETWORK (MANET). In MANET, the mobile nodes communicate with no fixed base station in an infrastructure less network [2]. The intermediate nodes act as the router which delivers the packets between two mobile nodes. Hence, MANET is a highly dynamic network and. In a dynamic MANET where every node participate in the process of routing, it is difficult to find out malicious node and hence, more vulnerable to attacks.

The routing protocol should posses the following properties [3], though all of these might not be possible to incorporate in a single solution.

- The routing protocol should be power-efficient.
- A routing protocol should be distributed to increase its reliability.
- Security should be the main factor of any routing protocol.
- A routing protocol should have good Quality of Service (QoS) [3].

- A hybrid routing protocol should be much more reactive than table driven to avoid overheads in the network.

To deliver the packets and the capability to handle dynamic connectivity are the most important concern for routing protocols. If there is an efficient path from source to destination then the protocol should use that path only to deliver the packets. If the connectivity of any two nodes is changed, then the routing protocol should be able to discover an alternate path if exists.

The different types of communications [4] used in mobile ad hoc networks are:

- Unicast: Unicast is a type of transmission in which only two nodes are exchanging the information. It is a one-to-one communication.
- Broadcasting: In Broadcast transmission, information is sent from just one node and received by all the other nodes connected to the network. One to all communication is called as broadcast.
- Multicasting: In this type of transmission, the information is sent to a set of nodes. It is a limited case of broadcasting. Multicasting lowers the communication cost for applications that transmits the same data to more recipients.

This paper is organized as follows. In Section 2, classifications of the routing protocols are given. In section 3, Properties of routing protocol is described. In Section 4, how AODV Routing Protocol works is explained and in section 5 the Attacks on AODV Routing Protocol is defined. In Section 6 Countermeasures for these Attacks are given and then Section 6 is the Conclusion of the paper.

7. CLASSIFICATION OF ROUTING PROTOCOL

Routing protocols are classified into two types based on their Properties.

- Proactive Routing Protocols

- Reactive Routing protocols.
- Hybrid routing protocol

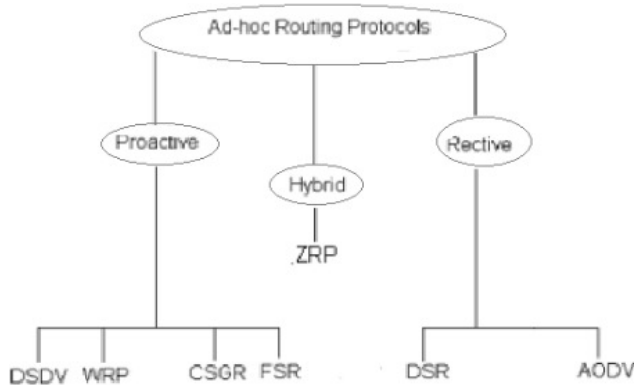


Fig. 1. Classification of Routing Protocol

A. Proactive Routing Protocols (Table driven)

In proactive or table-driven routing protocols [4], each node maintains up-to-date routes to every other node in the network by exchanging topological information among all the nodes in the network. Thus, when there is a need for a route to destination, such route information is available immediately. The areas in which the protocol differs are the number of necessary routing tables and the way these changes are broadcasted in network structure. The proactive protocols are not suitable for larger networks, as they need to maintain node entries for each and every node in the routing table. If the network topology changes very frequently, the cost of maintaining the network becomes be very high.

B. Reactive routing Protocols (On-demand)

With on-demand protocols [5], nodes do not keep correct routing information on all nodes at all the times. If any source node requires a route to the destination for which it does not have route information, it initiates a route discovery process which goes from one node to the other node until it reaches the destination or an intermediate node has a route to the destination. They do not need periodic transmission of topological information [4] of the network. The route discovery usually occurs by flooding the route request packets throughout the network.

C. Hybrid Routing Protocols

Hybrid routing protocol is the combination of both reactive and proactive protocol [4] which might yield better solution. These protocols organize the nodes in the groups and then providing different functionalities to the nodes inside and outside a group [5].

8. ADHOC ON-DEMAND DISTANCE VECTOR (AODV)

AODV is basically an improvement of DSDV [8]. But, AODV is an on-demand routing protocol. AODV is combination of both DSR and DSDV [4]. It minimizes the number of broadcasts by creating routes based on demand. When the source node wants to transmit any packet to a destination, it broadcasts a route request (RREQ) packet to its neighbors. The neighboring nodes in turn broadcast the packet to their neighbors and the process continues until the packet is reached by the destination. During the route request forwarding mechanism, intermediate nodes record the address of the neighbor from which the first copy of the broadcast packet is received. This record is stored in their route tables, which helps to establish the reverse path for the data packet. If repeated RREQs are later received by any node, then these packets are dropped. The reply is sent using the reverse path. The same process continues until the RREP message reaches to the source node. As the RREP is transmitted towards the source, all intermediate nodes set the forward route entries in their table for the transmission of data to the destination node. It may obtain multiple route to a destination so AODV uses destination sequence to determine up-to-date path to destination.

For route maintenance [3], If any intermediate node moves within a particular route, the neighbor of the drifted node can detect the link failure and sends a link failure notification to its neighbors and the process continues until the failure notification reaches the source node and the source can re-initiate the route discovery phase.

A. Route Discovery Mechanism in AODV:

If the source node “A” wants to initiate communication with destination node “E” as shown in the Figure 2 ,then it will make a connection between itself and the destination and will generate a route request message (RREQ). This message is then forwarded to the neighboring nodes, and the neighboring nodes will forward this control message to their neighboring nodes.

This process of finding destination node continues until the destination node is located itself or the node that has the fresh route to the destination. Once an intermediate node with enough fresh routes is located or destination node is located, they generate control message route reply message (RREP) and send it to the source node. When RREP reaches the source node, a route is established between the source node “A” and destination node “E”. Once the route is established between “A” and “E”, node “A” and “E” can communicate with each other. Figure 2 depicts the exchange of control messages between source node and destination node.

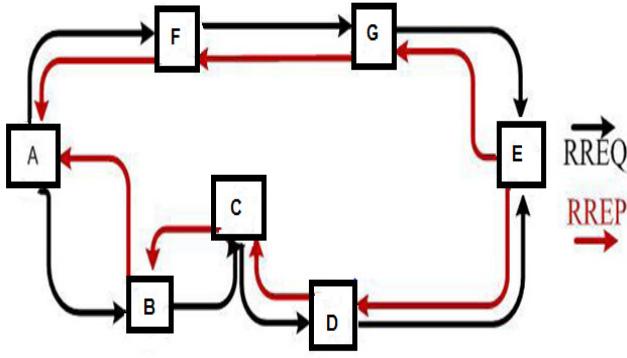


Fig. 2. Route Discovery Mechanism

B. Route Maintenance Mechanism

When there is a link down or a link breakage between destinations that causes one or more than one links unreachable from the source node or neighbor's nodes, then the RERR message is generated by the node and sent to the source node. If there is a route from "A" to "E" via "D", and if there is a link breakage "D" and "E", then the node "E" will generate and send the RERR message to the source node "A" informing the source node that there is a route error. The scheme is shown in the Figure 3 below.

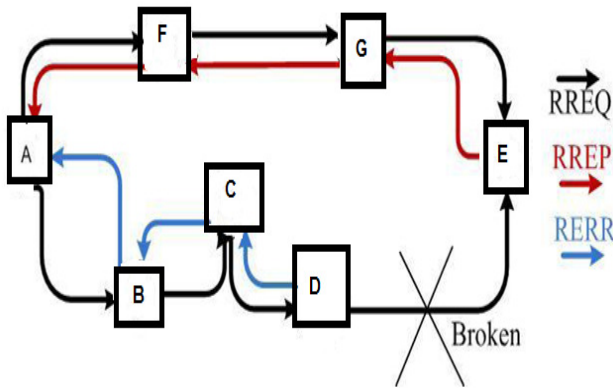


Fig. 3. Route Error Message

9. ATTACKS ON AODV ROUTING PROTOCOL

Attacks can be of 2 types: Passive Attacks and Active Attacks

- Passive Attack does not disturb the normal network operation while an active attack does it. Passive attacks are difficult to detect as there is no change in functionality of the network
- Active attack can be internal or external. Internal attacks can be carried out by nodes within the network while external attack can be carried out by nodes outside the network.[9]

In this section we will pay attention to the attacks, which are specifically applied and work on AODV routing protocol: Black hole attack, Byzantine attack and Wormhole attack. They will be presented as it follows:

- Black hole attack
- Rush attack
- Wormhole attack

A. Black hole attack

The black hole attack is an active insider attack with two properties [5]: first, the attacker does not forward any intercepted packet. Second, the node announces itself as the accurate route to reach to the destination node, even though the route is counterfeit. When the source node broadcasts the RREQ message to its neighbors, then the malicious node [14] also receives the forwarded RREQ message. The malicious node or the black hole node immediately sends an RREP message that contains the highest sequence number and this message is received by the source node as if it is having the best route to the destination. The source node considers the route with the black hole node.

In the following Figure 4 source node (node1) broadcasts a route request packet RREQ to its neighbors to find a route to the destination node 5. So, in this way, the node1 sends a RREQ packet to its neighbor node2, node3 and node4. We assume that node2 is the malicious node in the network and node4 have a valid route to the destination. Malicious node2 will instantly send a fake RREP to source node1 without even checking its routing table. So the fake RREP [10] reach faster to the node1 as compared to any other nodes in the network. At this time node1 accept the shorter route from the malicious node2 and rejects the other RREP packets and start the sending the data toward destination node through the path which has the malicious node. Source node1 believe that the data has been sent to destination node3, in fact, data has been intercepted by the malicious node2.

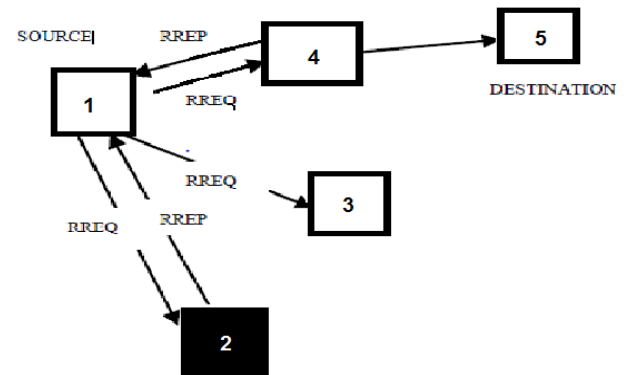


Fig. 4. Route Discovery with Black Hole Attack

B. Rushing Attack

In the rush attack, the attacker uses the higher transmission range or ignore the MAC [15] layer and/or routing layer delay to send the rushed routing packets (RREQ OR RREP) more quickly to the destination node in comparison to any other node in the network.

In Figure 5(a), it is the case of delay ignorance, node R1 is the rush attacker and sends the rushed RREQ more quickly to target node 3, in comparison to node 2. Node 3 discards the late forwarded legitimate RREQ [11] received from node 2 and forwards the first received R1 rushed RREQ to destination D. Consequently, D replies with RREP towards source via R1. As a result, source S transmit all the data packets through R1 and R1 discards all the data packets.

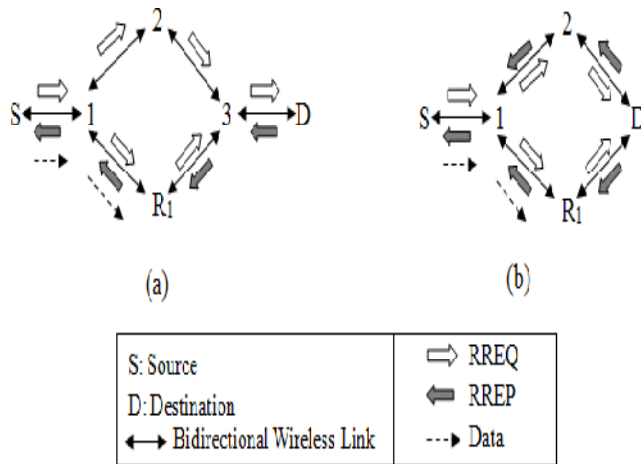


Fig. 5. Rushing attacker ignores MAC/routing layer delay[11]

In Figure 5 (b), node R1 is rush attacker and transmit rushed RREP faster to target Node 1 by ignoring delays, in comparison to node 2. Node1 drops the late received legitimate RREP [11] from node 2, and forwards the first received rushed RREP from R1 to S. As a result, S forwards all data towards R1 and R1 discards all those data.

Another type of rushing attacker sends rushed routing packets to target node using higher transmission range. In Figure 6 (a), node R2 is rushing attacker and sends rushed RREQ to target node 4 faster using higher transmission range, in comparison to node2. Node 4 drops all the late received legitimate RREQ form node 2 via node 3 and sends the first received rushed RREQ from R2 to destination D. Consequently, D replies with RREP towards source via R2 [11]. The wireless link between node4 and R2 is not bidirectional link as R2 is using higher transmission range. Now, the destination D forwarded RREP cannot be reached to R2 via node 4. As a result, the source node S cannot get the forwarded RREP and hence, no route will be discovered between S and D.

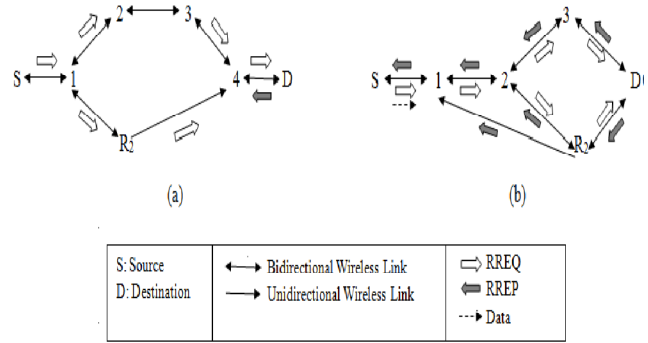


Fig. 6. Rushing attacker uses higher transmission range [10]

In Figure 6 (b), node R2 is rushing attacker which sends rushed RREP using higher transmission range. R2 sends rushed RREP to target node 1 faster, using the higher transmission range, in comparison to node3. Node1 drops all the late received legitimate RREP form node 3 via node 2 and forwards only the first received rushed RREP from R2 to source S. Consequently, S transmits all the data to destination D via R2. Now, the wireless links between R2 and node1 is not bidirectional link as R2 is using higher transmission range. So, When S forwarded RREP reaches node 1, it cannot be forwarded to R2 since of shorter transmission range of node 1. As a result, node D cannot get any data from S.

C. Wormhole Attack

In this type of attacks, the attacker interrupts the usual flow of routing packets. Wormhole attack can be done with one node or two or more nodes. But generally, two or more attackers are connected via a link called “wormhole link” [4]. The two malicious nodes in the network are located in the way that one near to the source node and another near to the destination node thus bypassing information [16] from source node to destination node and disrupting proper routing. They intercepts the packets at one end and replay them at the other end using private high speed network. The attacker tunnels the request packet RREQ directly to the destination node, without increasing the hop count value and thereby, prevents any other path from being discovered. Or it makes the tunneled packet arrives faster and with better metric value.

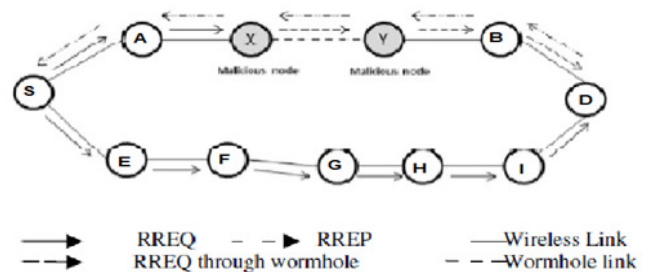


Fig. 7. Wormhole Attack

Wormhole attack associates two remote malicious nodes shown as X and Y in Figure 7 which are attached via a wormhole link and target to attack the source node S [4]. S broadcasts RREQ to find the route between source S and destination node D. Now, neighbors of S, A and E also broadcast the RREQ to their neighbors. Now, when the malicious node X receives RREQ forwarded by A, it tunnels the RREQ by the high-speed wormhole link to its partner Y. Malicious node Y forwards RREQ to the destination D via B. Thus, RREQ is forwarded via S-A-X-Y-B-D [17]. On the other hand, other RREQ packet is also forwarded through the path S-E-F-G-H-I-D. However, RREQ via X and Y reaches fast to D, as X and Y are connected via a high speed bus. Therefore, destination D discards all the RREQ packets that reach later and choose the path D-B-A-S [4] to send an RREP packet to the source node S. As a result, S chooses the route via X and Y to send data that to destination D.

10. COUNTERMEASURES AGAINST AODV ATTACKS

A. Counter Measures for Black Hole Attack

Deng et.al.[12] have proposed a solution against black hole attack by modifying the AODV protocol. In this approach, malicious nodes cannot advertise the route does not exist. In order to achieve this, each intermediate node includes the address of the next hop node in RREP packets. When the source node receives the RREP packet, it use the details of the next hop node and sends the request to the next hop node to verify the existence of the next hope node and the hop count value with the next hop. To confirm the route information, the next hop node of the neighbor node replies the further reply packet back to the source node. If the source does not receive the further reply, that means, the route contains the malicious nodes and the route is discarded from the routing table. However, this solution is vulnerable to cooperative black hole attacks. If both the next hop node and neighbor node are black hole nodes, the next hop node can response to the source node with false routing information.

In Ming – Yang Su et [13], they proposed that every Intrusion Detection System (IDS) node will execute a Anti Black Hole Mechanism (ABM) for estimating the suspicious value of a node by calculating the abnormal difference between RREQ and RREP message transmitted from the node. IDS will broadcast a block message if the suspicious value exceeds a thresh hold asking all nodes in the network to cooperatively isolate the malicious node with the help of this mechanism.

B. Counter Measures for Rushed Attack:

Rushing attacker sends rushed routing packets either by ignoring MAC/routing layer delay or by using higher transmission range. We have incorporated the trust concept

to isolate rushing attacker from discovered route in AODV Trusted On demand Routing (TOR) model [11]. Trust is belief or the trust level between trustor and trustee. Trustor (CT) evaluates the trust level for trustee (TE) on the basis of context dependent trust computation. Trustee (TE) is the node for which the trust level is being computed. The final trust computation is dependent on direct and indirect trust but under timing constraint. If CT evaluated Trust level for TE is not up to the threshold limit, then CT will not consider TE in route discovery process and discards the malicious TE routing packets.

C. TOR Model

In this work, AODV is extended with TOR model to establish route path between source and destination by discarding malicious nodes. The following TOR model (Figure 8) consists of three functional modules (Node Manager, Trust Module and Decision Manager) [11] along with the on-demand routing protocol, AODV.

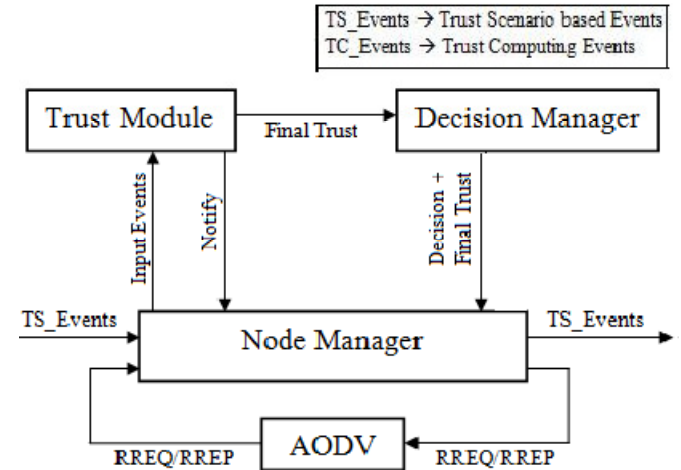


Fig. 8. TOR Model [11]

A. Trust Module

Trust Module of TOR model is responsible for trust value computation of TE. Trust Module consists of several components for computing different levels of trust values which are stored in repositories. Trust Module sends the computed final trust to decision Manager for taking belief-disbelief decision. On the other hand, Trust Module also stores the incoming notified final trust value for a TE.

B. Decision Manager

On the basis of received trust value from Trust Module, Decision Manager takes either accept or reject decision for TE. If the computed trust value is greater than 0.5, it takes accept decision for TE otherwise it takes reject decision. It then transmits this decision to Node Manager to take the

decision for considering or not considering TE in route discovery. It also forwards the final trust value to Node Manager for notifying other nodes in the network.

C. Countermeasures against WarmHole Attack

For detection and prevention of wormhole attacks, we are using a “Packet Leash” [18] mechanism in which the nodes obtain the authenticated symmetric key of every other node in the Mobile Adhoc network. By which, the receiver authenticates the information like destination and time from the received packet. To prevent the Network against wormhole attacks “Time of Flight” [17] method is used. This method calculates the roundtrip time of the message, to conclude whether the calculated distance is within the maximum possible communication range or not, acknowledgement is used to estimate the distance between the nodes based on this time. If there is an attacker in the network, packets end up travelling further.

Table 1 Listed Attack and Their countermeasures

ATTACKS	Method	Disadvantage
Black Hole Attack	<ul style="list-style-type: none"> • Include the address of the next hop in RREP packet. • ABM- Estimate the suspicious value by calculating the difference between RREQ and RREP 	<ul style="list-style-type: none"> • Not vulnerable to cooperative Attack. • Time delay and have to maintain extra database.
Rush Attack	<ul style="list-style-type: none"> • Introduce the TOR Model based on the Trust Value. 	<ul style="list-style-type: none"> • Difficult to implement and time delays.
Warmhole Attack	<ul style="list-style-type: none"> • “Time of Flight” technique is used and round trip journey time is calculated. 	<ul style="list-style-type: none"> • Extra database and time delays.

11. CONCLUSION

This paper includes the complex and dynamic behavior of Nodes in the AODV routing protocol in Mobile Adhoc network along with their attacks and respective countermeasures. We have described some common types of attacks in AODV: Black hole Attack, Rush Attack and Wormhole Attack. Every attack has some countermeasure to make the network more secured. Various Authors have proposed various solutions to prevent the Network from attack but every solution have their disadvantages. In this paper, we tried to inspect existing solutions for some common attacks occur on AODV. For future work, we are finding some points that can be further explored to protect

the MANET from ADHOC on Demand Distance Vector (AODV) attacks.

REFERENCES

- [1] C. Siva Ram Murthy and B. S Manoj, Ad Hoc Wireless Networks, Architecture And Protocols (Prentice Hall PTR, 2004)
- [2] Abhay Kumar Rai, Rajiv Ranjan Tewari, Saurabh Kant Upadhyay,” Different Types of Attacks on Integrated MANET-InternetCommunication” proceeding of *International Journal of Computer Science and Security (IJCSS) Volume (4): Issue (3)*.
- [3] G.Vijaya Kumar, Y.Vasudeva Reddyr , Dr.M.Nagendra,” Current Research Work on Routing Protocols for MANET: A Literature Survey” proceeding of *(IJCE) International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 706-713*
- [4] Deepa.S, Dr. D.M Kadhar Nawaz,” A Study on the Behavior of MANET Routing Protocols with Varying Densities and Dynamic Mobility Patterns” proceeding of *IJCA Special Issue on “Mobile ADHOC Networks” MANETs, 2010*.
- [5] Amol A. Bhosle, Tushar P. Thosar and Snehal Mehatre, “Black-Hole and Wormhole Attack in Routing Protocol AODV in MANET ”proceeding of *International Journal of Computer Science, Engineering and Applications (IJCEA) Vol.2, No.1, February 2012*.
- [6] Rajni Sharma1, Alisha Saini2,” A Study of Various Security Attacks and their Countermeasures in MANET ” proceeding of *International Journal of Advanced Research in Computer Science and Research*.
- [7] Guangyu Pei, Mario Gerla, Tsu-Wei Chen,” Fisheye State Routing in Mobile Ad Hoc Networks” proceeding of *NSF in part of DARPA*.
- [8] Vidya Shree. P, Sophia Reena.G., “A SURVEY OF VARIOUS ROUTING PROTOCOLS IN MOBILE ADHOC NETWORKS (MANET)” in proceeding of *International Journal of Computer Science & Engineering Technology (IJCSET)*.
- [9] G.S. Mamatha, Dr. S.C. Sharma, “Network Layer Attacks and Defense Mechanisms in MANETS- A Survey” in proceeding of *International Journal of Computer Applications (0975 – 8887) Volume 9– No.9, November 2010 12*
- [10] Govind Sharma, Manish Gupta,“ Black Hole Detection in MANET Using AODV Routing Protocol” in proceeding of *International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-6, January 2012 297*.
- [11] Swarnali Hazra and S.K.Setua,“RUSHING ATTACK DEFENDING CONTEXT AWARE TRUSTED AODV IN ADHOC NETWORK” in proceeding of *International Journal of Security, Privacy and Trust Management (IJSPTM), Vol. 1, No 3/4, August 2012 DOI : 10.5121/ijspmt.2012.1305 53*.
- [12] Madhusudhananagakumar KS, G.Aghila “A Survey on Black Hole Attacks on AODV Protocol in MANET ”in proceeding of *International Journal of Computer Applications (0975 – 8887) Volume 34– No.7, November 2011 23*.
- [13] Varsha Patidar1, Rakesh Verma2,” Black Hole Attack and its Counter Measures in AODV Routing Protocol” in proceeding

of *International Journal of Computational Engineering Research* (*ijceronline.com*) Vol. 2 Issue. 5.

- [14] Pinki Tanwar, Shweta, "A SURVEY ON BEHAVIOUR OF BLACKHOLE IN MANETS" proceedings of *IJRIM Volume 1, Issue 4 (August, 2011)*.
- [15] Moitreyee Dasgupta, S. Choudhury, N. Chaki, "Routing Misbehavior in Ad Hoc Network" in proceedings of *2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 18*.
- [16] Sana ul Haq, Faisal B Hussain "OUT-OF-BAND WORMHOLE ATTACK DETECTION IN MANETS" in proceedings of 9th Australian Information Security Management Conference, Edith Cowan University, Perth Western Australia, 5th -7th December, 2011
- [17] Rutvij H. Jhaveri, Ashish D. Patel, Jatin D. Parmar, Bhavin I. Shah, "MANET Routing Protocols and Wormhole Attack against AODV" in proceedings of *IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.4, April 2010* 12 Manuscript received April 5, 2010 Manuscript revised April 20, 2010.
- [18] T. Sakthivel, R. M. Chandrasekaran, "Detection and Prevention of Wormhole Attacks in MANETs using Path Tracing Approach" in proceedings of *European Journal of Scientific Research ISSN 1450-216X Vol.76 No.2 (2012), pp.240-252*

Network Layer Attacks and its Countermeasure in Mobile Ad hoc Networks

Avinka Baweja¹, Kushagra Agrawal², Sanjeev Rana³, Rajneesh Kumar Gujral⁴

^{1, 2, 4}Department of Computer Science & Engineering

³Department of Information Technology

M.M. Engineering College, M.M. University, Mullana, Ambala (133207)

¹avinkabaweja@gmail.com, ²agrawal_kushagra@rediffmail.com, ³dr.sanjeevrana@yahoo.com,

⁴dr Rajneeshgujral@yahoo.com

Abstract: Security in mobile ad hoc networks is difficult to achieve, because of the vulnerability of wireless links, limited physical protection of nodes, dynamically change of topology, the absence of a certification authority, and the lack of a centralized management point. Initial studies on mobile ad hoc networks (MANETs) aimed at proposing protocols for fundamental problems, such as routing. These protocols fully trust all nodes and do not consider security as a major aspect. They are consequently vulnerable to attacks and misbehavior. In recent years, researchers focused on security problems in MANETs, and proposed mechanisms to secure protocols and applications. In this paper surveys some attacks on network layer and present some existing countermeasure to tackle them.

Keywords: MANETs, AODV, security

1. INTRODUCTION

An AD-HOC network is a collection of mobile nodes and wireless communication network is used to connect these mobile nodes. This type of network is known as MOBILE ADHOC NETWORK (MANET). Each device in a MANET moves independently. MANET is an infrastructure less network with no fixed BS for communication. Intermediate mobile nodes act as router to deliver the packets between two nodes. So, MANET is a highly dynamic network and hence more vulnerable to attack. In a dynamic MANET where every node taking participate in the routing process, it is difficult to find out malicious node.

There are a variety of attacks on network layer which may harm the network. Like malicious routing attack may harm operation of the routing protocols. There is a wide variety of attacks that weakness of network. For example, routing messages are an important component for mobile communications, as each packet need to be passed quickly through intermediate nodes. Malicious routing attacks can spoil the routing discovery or maintenance phase by not following the specifications of the routing protocols. There are also attacks that target some particular routing protocols, such as DSR, or AODV[1,2,3].

Some sophisticated attacks have been identified are Black hole, Worm hole, Flooding and Byzantine attack. This paper is organized as follows. In Section II, classifications of the security attacks are given. In section III, routing protocols in network layer. In Section IV, proposed attacks, Section V include their countermeasures for the network layer and section VI is the conclusion of the paper.

2. CLASSIFICATION OF SECURITY ATTACKS

The attacks in MANETS are classified into two major categories,

- passive attacks
- active attacks

Passive attacks are those, launched by the adversaries just to look the data exchanged in the network. These adversaries don't disturb the operation of the network. It becomes very difficult to identify such attack as the network itself does not affected and they can reduced by using powerful encryption techniques.

Active attack tries to alter or destroy the information that is being exchanged, thereby disturbing the normal functionality of the network.[3]

In figure 1 the classification of Network security attacks against MANETs is presented. Passive attacks can be listed as eavesdropping, traffic analysis, and traffic monitoring. Active attacks include wormhole, black hole, gray hole, information disclosure, resource consumption, routing attacks and others include jamming, impersonating, modification, denial of service (DoS), and message replay.[3]

3. ROUTING PROTOCOL IN NETWORK LAYER

Table Driven Routing Protocols (Proactive): In proactive or table-driven routing protocols, each node maintains a routing table and continuously maintains up-to-date routes to

every other node in the network by exchanging the information among all the nodes in the network. So, when there is a need for a route to destination, these routing tables are available to tell the route to destination immediately. As these tables need to maintain node entries for each and every

node in the routing table and due to which proactive protocols are not suitable for larger networks. If the network topology changes too frequently, the maintenance the network will be very high.

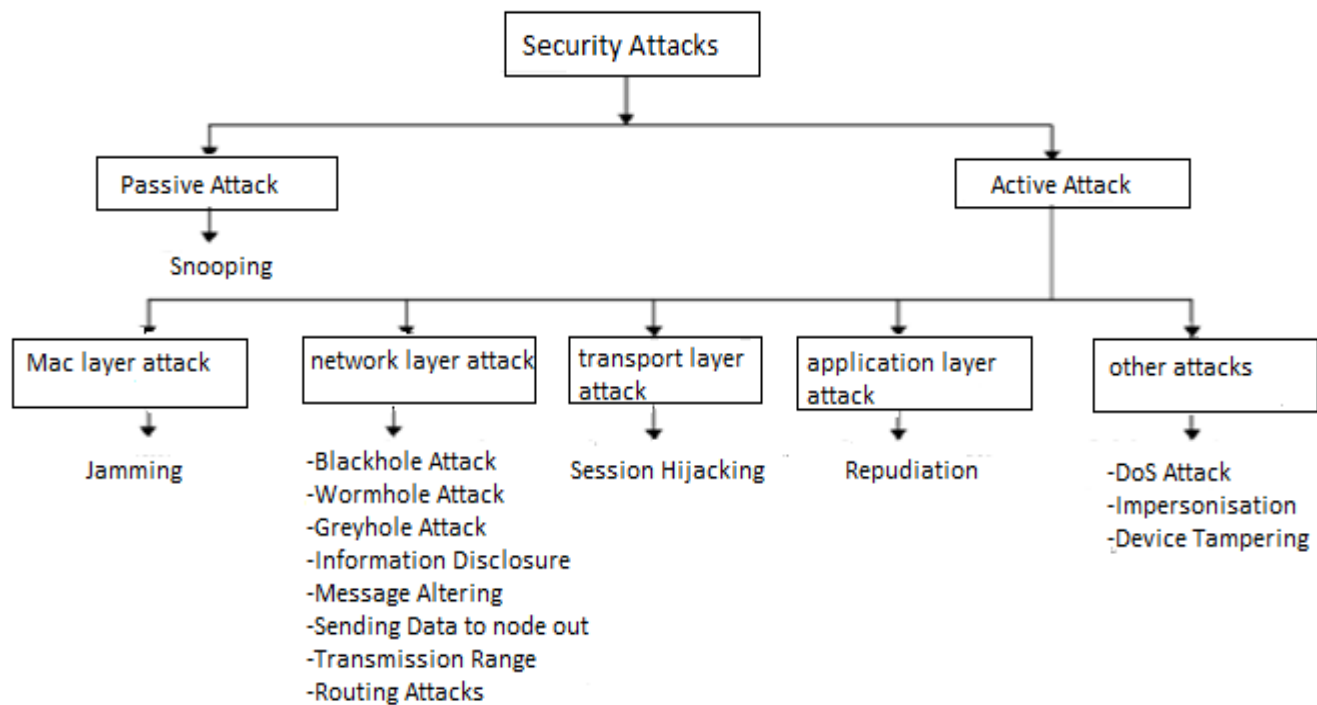


Fig. 1. Classification of Security Attacks [3].

On-Demand routing Protocols (Reactive): With on-demand routing protocols, if a source node requires a route to the destination and it does not have the route to destination, so it initiates a route discovery process which goes from one node to the other until it reaches to the destination or an intermediate node that has a route to the destination. This protocol does not need periodic exchange of routing information. The route discovery process usually done by flooding the route request packets to its neighboring nodes.

Hybrid Routing Protocols: Since, reactive or proactive feature of a particular routing protocol is not enough; So we introduce a better routing protocol which is a mixture of both reactive routing protocol and proactive routing protocol that yield a better solution. Hence, in the recent years, several hybrid protocols are also proposed.

A. Proactive Routing Protocols

1) Dynamic Destination-Sequenced Distance-Vector Routing Protocol (DSDV): In this routing protocol, each mobile node in the network keeps a routing table. Each of the routing

table includes all available destinations and the number of hops to reach that destination. Each entry in the routing table has a sequence number. If a Link is present then sequence number will be even otherwise odd number will be used. This number is generated by the destination, and the sender node should have to send out the next update with this number. The Periodic transmissions of updated routing tables help to maintain the topology information in the network. If there is any new and significant change in the network then the updates will be sent out immediately to the neighbors. So, the routing information updates may be periodic or when any topology change occurs. DSDV protocol each mobile node in the network will send its routing table to its current neighbors. This is possible either by broadcasting or by multicasting. By the advertisements, the neighboring nodes can know whether any change has occurred in the network due to the movements of nodes. The routing updates could be sent in two ways: one is called a “full dump” and another is “incremental.”[10] In full dump, the entire routing table is transmitted to the neighbors, when change occurs in the topology. But in case of incremental update only the entries that are updated due to changes are sent.

2) *Wireless Routing Protocol (WRP)*: WRP uses an enhanced version of the distance-vector routing protocol, which uses the Bellman-Ford algorithm to calculate shortest paths using information regarding the length and second-to-last hop of the shortest path to each destination [10,11]. To overcome the count-to-infinity problem, it uses a unique method of maintaining information of the shortest distance to every destination node in the network. It is different from DSDV in table maintenance and Table update procedures. In Wireless Routing Protocol, each node maintains four things: 1. A distance table 2. A routing table 3. A link-cost table 4. A message retransmission list (MRL). In WRP, nodes check for the existence of their neighbors by sending the periodic update messages to the neighbors. The recipient nodes of update message should send the acknowledgments. If a node is not sending any packets, then the nodes in the response list should send an idle Hello message to ensure connectivity. A node can decide whether to update its routing table after receiving an update message from a neighbor and always it looks for a better path using the new information. If a node gets a better path, it sends the copy of its routing table information back to the original nodes so that they can update their tables. After receiving the acknowledgment, the original node updates its MRL. Thus, it avoids count to infinity problem.

3) *Cluster Gateway Switch Routing Protocol (CGSR)*: CGSR uses a clustered mobile wireless network. In this type of Routing Protocol the network is divided into separate but interrelated groups, one of the nodes is chosen as the cluster head using least cluster change (LCC) algorithm. By separating network into clusters, this protocol achieves a distributed processing mechanism in the network. But the drawback of this protocol is that, frequent change or selection of cluster heads may affect the routing performance of the protocol by consuming resources. CGSR uses DSDV protocol for routing. This protocol modifies DSDV by using a hierarchical cluster-head-gateway routing approach for routing. The nodes that are within the communication ranges of two or more cluster heads are Gateway nodes. The communication between the clusters takes place through this node. A packet sent by a node in a cluster is first sent to its cluster head, and then the packet is sent from the cluster head to a gateway to another cluster head, and so on until the cluster head of the destination node is reached. The packet is then transmitted to the destination node.

4) *Fisheye State Routing (FSR)*: FSR uses a special structure of the network called the "fisheye." This protocol basically reduces the traffic for transmitting the update messages. The basic idea is that each update message contains the information about its neighboring node instead of containing information about all nodes in the update message. Therefore, each node has accurate and exact information about its own neighboring nodes. The reduction of update

message size is obtained by using different exchange periods for different entries in the table [10,13]. The node entries in the message amount of entries are reduced which reduces the message size. This Protocol can give the best path to distant path to a destination with the fact that the route will be progressively more accurate as the packet gets closer to its destination.

B. Proposed on demand Routing Protocol

1) *Dynamic Source Routing (DSR)*: The basic idea of DSR is that, it uses the concept of source routing, which means the sender knows the complete hop-by-hop route to the destination. In this protocol, all the mobile nodes are required to maintain route caches which contain the routes to other nodes. The route cache is updated only when any new route is maintained/updated for a particular entry in the route cache. The data packet carries the source route in the packet header. Routing in DSR is done in two phases: route discovery and route maintenance. When a source node wants to send a packet to a destination, it first checks its route cache to determine whether its cache already contains any route to the destination or not. If there is already an entry for that destination, the source uses that route to send the packet. If not, then the source node broadcasts a route request packet which includes the destination address, source address, and a unique request ID. Each intermediate node checks whether the route is available or not. If the intermediate node does not know the route to destination, it adds its own address to the route request packet and forwards the packet and to other nodes eventually this reaches the destination. [11,12] A node processes the route request packet only if it is not previously processed that packet. A route reply is generated by the destination or by any of the intermediate nodes which know the route to destination.

Another Phase is Route Maintenance which is done by using the route error packet (RERR) and acknowledgements. Route error packets are generated by a node if there is any Link break occurs or any other error in the route. When a route error packet is received by node, the hop in error is removed from the route cache.

2) *Ad Hoc On-Demand Distance Vector Routing (AODV)*: AODV is basically an improved DSDV. But, AODV is a reactive routing protocol instead of proactive. It minimizes the number of broadcast messages by creating routes based on demand, which is not happen in DSDV. When any source node wants to send a packet to a destination, it broadcasts a route request (RREQ) packet. The neighboring nodes in turn broadcast the packet to their neighbors and the process continues until the packet reaches the destination. During the process of forwarding the route request packet, intermediate nodes record the address of the neighbor from which the first copy of the broadcast packet is received. This record is stored in their routing tables, which helps to establish a

reverse path. If additional copies of the same RREQ are later received, these packets are discarded. The reply is sent using the reverse path, the same process continues until the RREP message reaches to the source node. As the RREP is propagated back towards the source, all intermediate node sets the forward route entries in their table for the transmission of data packages to the destination node. It may obtain multiple route to a destination. so AODV uses destination sequence to determine up-to-date path to destination. [10,11,12]

For route maintenance, when a source node moves, the route discovery process will restart. If any intermediate node moves, the neighbor of the moved node can detect the link failure and sends a link failure notification to its neighbor. This process continues until the failure notification reaches the source node. On the receiving information Link failure information, the source may re-initiate the route discovery phase.

C. Proposed Hybrid Routing Protocol

1) *Zone routing Protocol*: Zone Routing Protocol or ZRP is the first hybrid routing protocol have the features of both proactive and reactive routing component. ZRP was proposed to reduce the control overhead of proactive routing protocols and decrease the latency caused by route discovery in reactive routing protocols [10]. The Zone Routing Protocol, as the name implies, is based on the concept of zones. Every node has its own separate routing zone and the zones of neighboring nodes overlap with others. To create zones in the network, a node first has to know about its neighbors. A neighbor is defined as a node with which direct communication can be established, and that is; one hop distance from a node. A route to a destination within the local zone can be established from the source's proactive route cache. For routes that are not in local zone, route discovery happens according to reactive protocol by sending the route request to the peripheral nodes of its zone. Each peripheral node checks its local zone for the destination. If the destination is not there, then the peripheral node adds its own address to the route request packet and forwards the packet to its own peripheral node. If the destination is a member of the local zone, it sends a route reply on the reverse path back to the source.

4. ATTACKS ON MANET NETWORK LAYER

In this section we will pay attention to the attacks, which are specifically applied and work against MANET network layer: flooding attack, Black hole attack, Byzantine attack and Wormhole attack.

They will be presented as it follows:

- Flooding attack

- Black hole attack
- Wormhole attack
- Byzantine attack

A. Flooding attack

There are different types of flooding attacks, which may disrupt the routing discovery or the maintenance phase of routing within MANET. Basically, In a flooding attack a malicious node/an attacker aims the exhaustion of the network resources (e.g. network bandwidth) as well as consuming the resources of an authentic network user (e.g. computational and battery power)[6]. Furthermore an attacker can influence the network performance, by obstructing the proper execution of routing protocol. By RREQ flooding (or routing table overflow) is possible for an attacker to send multiple RREQs to recipient node which actually doesn't exist. It means the malicious node represents false routes to all authentic nodes within the network, which makes it impossible to create new and actual routes which causes routing table overflow by the authentic users. The flood of RREQs all over the network leads to consumption of the battery power and the network bandwidth which decreases the performance of the network.

B. Black hole attack

The black hole attack is an active insider attack with two properties [15]: first, the attacker does not forward any intercepted packet. Second, the node announces itself as the accurate route to reach to the destination node, even though the route is counterfeit

This attack takes 2 steps to proceed:

- 1) A malicious node modifies the network topology to create "environment" for the attack. It presents itself as a genuine node within the network, and gives a genuine route to destination, aiming to capture the information.
- 2) In the second step of Black hole attack the malicious node consumes the intercepted data packages; it simply captures the information and does not forward it to the end user.

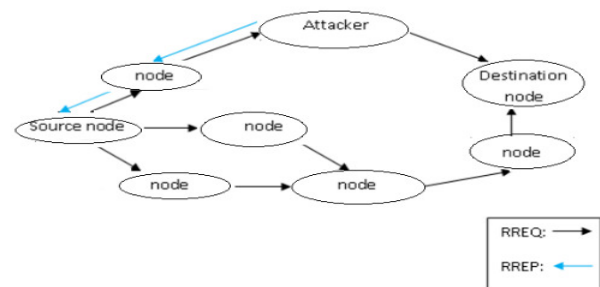


Fig. 2. How Black hole attack works

In the following paragraph, let's take a closer look at the Black hole attack showed in Figure 2. The source node sends RREQs to all nodes in the network to find out the possible authentic routes to the destination. As the attacker receives the RREQ sent by the source node it forwards it to the destination node and send a RREP back to the source node and present itself as a genuine node by giving genuine route. After picking up the data from the source node for the transfer of the data as an authentic user within MANET, it will receives the information by intercepting the data flow and does not forward it to the destination node.

C. Wormhole attack

The wormhole attack is one of the most effective and hard attacks, which can be executed within MANET. In this type of Attack, two collaborating attackers establish a link, called wormhole link by using private high speed network. With this wormhole link, one of the attacker can capture the data packet in between the data flow and send them to second attacker via wormhole link and replays them. The Worm hole node results in denial of service as it discards all the data packets instead of forwarding[14].

In the following paragraph, let's take a closer look to the wormhole attack, as shown in figure 3: The target node sends RREQs all over the network to find out the possible authentic routes. As the attacker 1 receives the RREQ sent by the target node and forwards it to the attacker 2 over the wormhole link between them. As the attacker 2 receives the RREQ, transmit it to the destination node. The destination node on its part sends a RREP back to the target node over the wormhole link between the two attackers. In order to present itself as a genuine route, the attackers forward the RREP to the source node. After getting the RREP from the destination node, the data is transmitted by source node over the same false route within MANET, now the attackers can intercept the data flow, i.e. receive the information and does not forward it to the end user (destination node), or selectively forward data packages.

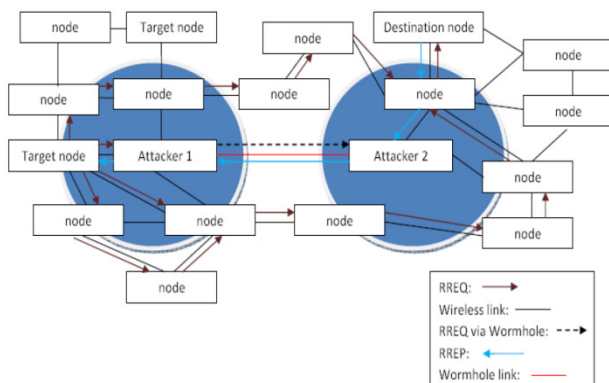


Fig. 3. How worm hole attack works[6].

D. Byzantine Attack

In this type of attack the compromised or malicious nodes tries to create routing loops or routing of the data packets on the non-optimal routes or selectively drop packets. This kind of failures is not easy to identify, since the network seems to be operating very normally in the view of the user [3].

Or in more general terms : A compromised intermediate node works alone, or a set of compromised intermediate nodes works together and carry out attacks such as creating routing loops, forwarding packets through non-optimal paths, or selectively dropping packets, which results in disruption or degradation of the routing services [8].

5. COUNTERMEASURES

A. Countermeasure against the flooding attack

Every authentic node can monitor and compute the evaluation of its neighbor's RREQ where the RREQ limit is predefined. In case of outmatching of RREQ's limit, the specific node with its ID comes in a list, called Blacklist and the authenticated nodes will not send and receive any RREQ packet to this node.

Another countermeasure is: A novel scheme which is based on a firewall. This firewall can filter out the packets sent by attacker from the packets sent by authenticated users based on the marking value on the packet, and thus filter out most of the attack packets. This scheme is very effective and has a very low deployment cost as compared to other packet-marking based solutions, In the implementation of this scheme the server will generate encrypted marking for secure transmission of packets. The scheme allows the firewall to Detect and prevent the Denial of service attacks from the first packet itself [4].

B. Countermeasure against Blackhole attack

As in flooding attack, neighbor's nodes could detect the sequence of the falsified RREQ or RREP messages and put the malicious node in their blacklists, stop sending and receiving data from it. So, to minimize the risk of being expose, the malicious node work more efficiently by not entirely intercepting between 2 interacting nodes, but can selectively forward packets. But this scheme has some disadvantage the attacker can still modify some packets sent from a particular node not from all.

Another counter measure is: Security-aware ad hoc routing protocol (SAR) can be used to prevent the nodes against black hole attacks. In SAR, a security metric is added into the RREQ packet, and a different route discovery procedure is used. All nodes receives RREQ packet with a particular security metric in that packet. At intermediate nodes, if the security metric is satisfied, the node will process the RREQ

packet, and it will forward to its neighbors. Otherwise, the RREQ is dropped. If an source to destination path with the required security metric found, the destination will generate a RREP packet with that security metric. If the destination node fails to find a route with the required security metric, it sends a notification to the sender to adjust the security metric to find a route.

C. Countermeasure against the Wormhole attack

Wormhole Attack Prevention (WAP) without using specialized hardware is proposed to prevent the wormhole attack [5]. The WAP not only detects the false route but it also use preventive measures against wormhole nodes, so that they will not reappear again during the route discovery phase.

Another countermeasure is: Packet leashes can be used to detect wormhole attacks. A leash is the information added into a packet to tell its transmission distance. A leash sets the lifetime of a packet, which restricts its transmission distance. A sender includes the transmission time and destination in the message. The receiver checks whether the packet has traveled the distance between the sender and itself within the specified time. Packet leashes require highly synchronized clocks and precise knowledge of location.

D. Countermeasure against the Byzantine attack

A secure on-demand MANET routing protocol, named Robust Source Routing (RSR) [9] is proposed as countermeasure of Byzantine attacks.

In Robust Source Routing, an data origin authentication services and integrity checks are performed by which RSR is able to fight against the malicious nodes which selectively drop the packets or modify the packets [3].

Table 1: List of Attacks and its countermeasures

Network layer Attacks	How Attacker works	Countermeasures
Flooding Attack	Malicious Node aims the exhaustion of network resources as well as the consumption of resources by representing false routes to authentic nodes.	<ul style="list-style-type: none"> By monitoring neighbors RREQ. By Novel Based Scheme (which is based on Firewall).
Blackhole Attack	Malicious node presents itself as a genuine node and simply receives the information.	<ul style="list-style-type: none"> By monitoring neighbor's RREQ and RREP. By Security Aware Routing Protocol (A security metric added to RREQ packet).
Wormhole Attack	Two or more collaborating Attackers establish a wormhole link and captures the data packets.	<ul style="list-style-type: none"> By using Wormhole Attack Prevention Scheme (WAP). By using Leash to the data packet(which tells the lifetime of the packet)
Byzantine Attack	Malicious Node may create Loops or may route the packets to non-optimal paths.	<ul style="list-style-type: none"> By using Robust Source Routing (RSR).

6. CONCLUSION

This paper pays attention to the complex infrastructure of Mobile Ad hoc Network attacks and its countermeasure. The theoretical fundamentals of the different types of security attacks are represented to give an overview of the attacks on Network Layer. All the Network Layer Protocols are described on which attacks may occur. Afterwards, some common attacks like Flooding Attack, Blackhole attack, Wormhole Attack and Byzantine Attack is described, which can occur in MANET. Every attack, has some countermeasure to make the network more secured. Various Authors have proposed various solutions to prevent the Network from attack but every solution have their disadvantages. In this paper, we tried to inspect existing solutions for some common attacks occur on Network Layer. In future, we will try to find a common solution, which will encounter all attacks present in network layer.

REFERENCES

- [1] C.Sivaram Murthy and B.S. Manoj,"ADHOC Wireless Networks: Architecture and Protocols", Prentice Hall PTR, 2004.
- [2] Pradeep Rai, Shubha Singh," A review of MANET's Security Aspects and challenges", Proceedings of "IJCA special issue on Mobile Adhoc Networks, 2010.
- [3] G.S. Mamatha and Dr. S.C. Sharma, "Network Layer Attacks and Defense Mechanisams in Manets-A Survey", Proceedings of "International Journal of Computer Applications, November 2010.
- [4] Abhay Kumar Rai, Rajiv Ranjan Tewari & Saurabh Kant Upadhyay," Different Types of Attacks on Integrated MANET-Internet Communication".
- [5] Pradeep M Jawandhiya, Mangesh M Ghonge, Dr. M S Ali," Countermeasures of Network Layer Attacks in MANET", Proceeding of IJCA Special Issue on "Network Security and Cryptography" NSC, 2011.
- [6] Zdravko Danailov," Attacks on Mobile Ad hoc Networks", A Seminar Report
- [7] Rusha Nandy, Debduutta Barman Roy," Study of Various Attacks in MANET and Elaborative Discussion of Rushing Attack on DSR with Clustering Scheme", Proceeding of IJAN and applications ,2011.
- [8] Bing Wu, Jianmin Chen, Jie Wu, Mihaela Cardei," A Survey on Attacks and Countermeasure in Mobile Ad Hoc Networks",Proceeding of Y. Xiao, X. Shen, and D.- Z. Du (Eds.) pp. -- -- 2006 Springer.
- [9] Claude Cr'epeau, Carlton R. Davis and Muthucumaru Maheswaran, "A secure MANET routing protocol with resilience against byzantine behaviors of malicious or selfish nodes", 21st International Conference on Advanced Information Networking and Applications Workshops 2007.
- [10] G.Vijaya Kumar, Y.Vasudeva Reddy and Dr.M.Nagendra." Current Research Work on Routing Protocols for MANET: A Literature Survey". Proceeding of (IJCSE) International Journal on Computer Science and Engineering 2010.

- [11] Vidya Shree.P, Sophia Reena.G., "A Survey of various routing Protocol in MANET",proceeding of International Journal of Computer Science & Engineering Technology (IJCSET).
- [12] Deepa.S, Dr. D.M Kadhar Nawaz," A Study on the Behavior of MANET Routing Protocols with Varying Densities and Dynamic Mobility Patterns", proceeding of IJCA Special Issue on "Mobile Ad-hoc Networks MANETs, 2010.
- [13] Guangyu Pei, Mario Gerla and Tsu-Wei Chen," Fisheye State Routing in Mobile Ad Hoc Networks ".
- [14] T. Sakthivel, R. M. Chandrasekaran," Detection and Prevention of Wormhole Attacks in MANETs using Path Tracing Approach" in proceedings of European Journal of Scientific Research ISSN 1450-216X Vol.76 No.2 (2012), pp.240-252.
- [15] Amol A. Bhosle, Tushar P. Thosar and Snehal Mehatre, "Black-Hole and Wormhole Attack in Routing Protocol AODV in MANET "proceeding of International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.1, February 2012.

Permission-Based Security Models and its Application to Android System

Ahmad Talha Siddiqui¹, Munesh Chandra Trivedi²

¹Research Scholar, ²Associate Professor & Head

¹IFTM University, ²Department of Computer Science & Engineering
¹ahmadtalha2007@gmail.com, ²DIT, School of Engineering, Greater Noida

Abstract: The market for smart phones has been booming in the past few years. There are now over 400,000 applications on the Android market. Over 10 billion Android applications have been downloaded from the Android market. Due to the Android popularity, there are now a large number of malicious vendors targeting the platform. Many honest end users are being successfully hacked on a regular basis. In this work, a cloud based reputation security model has been proposed as a solution which greatly mitigates the malicious attacks targeting the Android market. Our security solution takes advantage of the fact that each application in the android platform is assigned a unique user id (UID). Our solution stores the reputation of Android applications in an anti-malware providers' cloud (AM Cloud). The experimental results witness that the proposed model could well identify the reputation index of a given application and hence its potential of being risky or not.

Keywords: Smart phones; Android OS; Reputation based security; Inter Process Communication

1. INTRODUCTION

Access control lists (ACLs) and permission-based security models allow administrators and operating systems to restrict actions on specific resources. In practice, designing and configuring ACLs (particularly those with a large number of configuration parameters) is a complicated task. More specifically, reaching a balance between the detailed expressiveness of permissions and the usability of the system is not trivial, especially when a system will be used by novices and experts alike. One of the main problems with ACLs and permission models in general is that they are typically not designed by the users who will ultimately use the system, but rather by developers or administrators who may not always for see all possible use cases. While some argue that the problem with these permission-based systems is that they are not designed with usability in mind [11], we believe that in addition to the usability concerns, there is not a clear understanding of how these systems are used in practice, leading security experts to blindly attempt to make them better without knowing where to start. While there are many widely deployed systems which use permissions, we focus on the empirical analysis of the permission model included in Android OS [1]. Android is a newcomer to the

smart phone industry and in just a few years of existence has managed to obtain significant media attention, market share, and developer base. Android uses ACLs extensively to mediate inter-process communication (IPC) and to control access to special functionality on the device (e.g., GPS receiver, text messages, vibrator, etc.). Android developers must request permission to use these special features in a standard format which is parsed at install time. The OS is then responsible for allowing or denying use of specific resources at run time. The permission model used in Android has many advantages and can be effective in preventing malware while also informing users what applications are capable of doing once installed.

The main objectives of our empirical analysis are: (1) to investigate how the permission-based system in Android is used in practice (e.g., whether the design expectations meet the real-world usage characteristics and (2) to identify the strengths and limitations of the current implementation. We believe such analysis can reveal interesting usage patterns, particularly when the permission-based system is being used by a wide spectrum of users with varying degrees of expertise.

2. BACKGROUND

Access control systems have existed for a long time [17]. In its basic form, a security system based on access control lists allows a subject to perform an action (e.g., read, write, run) on an object (e.g., a file) only if the subject has been assigned the necessary permissions. Permissions are usually defined ahead of time by an administrator or the object's owner. Basic file system permissions on POSIX-compliant systems [12] are the traditional example of ACL-based security since objects – in this case, files can be read, written or executed either by the owner of the file, users in the same group as the owner, and/or everyone else. More sophisticated ACL-based systems allow the specification of a complex policy to control more parameters of how an object can be accessed. We use the term permission-based security to refer to a subset of ACL-based systems in which the action doesn't change (i.e., there is only one possible action to allow or deny on an object). This would be similar

to having multiple ACLs per object, where each ACL only restricts access to one action. We note that reducing the allowable actions to one does not necessarily make the system easier to understand or configure. For example, in the Android permission model, developers implement finer level granularity by defining separate permissions for read and write actions.

2.1 Permission-Based Security Examples

An example of a permission-based security model is Google's Android OS for mobile devices. Android requires that developers declare in a manifest a list of permissions which the user must accept prior to installing an application. Android uses this permission model to restrict access to advanced or dangerous functionality on the device [14]. The user decides whether or not to allow an application to be installed based on the list of permissions included by the developer. Similar to Android OS, the Google Chrome web browser uses a permission-based architecture in its extension system [4]. Extension developers create a manifest where specific functionality (e.g., reading bookmarks, opening tabs, contacting specific domains) required by the extension can be requested. The manifest is read at extension install time to better inform the user of what the extension is capable of doing, and reduce the privileges that extensions are given [10]. In contrast, Firefox extensions, which do not have this permission architecture, run all extension code with the same OS-level privileges as the browser itself. A third example of a currently deployed permission-based architecture is the Blackberry platform from Research in Motion (RIM). Blackberry applications written in Java must be cryptographically signed in order to gain access to advanced functionality (known as Blackberry APIs with controlled access) such as reading phone logs, making phone calls or modifying system settings [3].

2.2 Related Work

Enck et al. [13] describe the design and implementation of a framework to detect potentially malicious applications based on permissions requested by Android applications. The framework reads the declared permissions of an application at install time and compares it against a set of rules deemed to represent dangerous behaviour. For example, an application that requests access to reading phone state, record audio from the microphone, and access to the Internet could send recorded phone conversations to a remote location. The framework enables applications that don't declare (known) dangerous permission combinations to be installed automatically, and defers the authorization to install applications that do to the user.

Ontang et al. [18] present a fine-grained access control policy infrastructure for protecting applications. Their proposal extends the current Android permission model by

allowing permission statements to express more detail. For example, rather than simply allowing an application to send IPC messages to another based on permission labels, context can be added to specify requirements for configurations or software versions. The authors highlight that there are real-world use cases for a more complex policy language, particularly because untrusted third-party applications frequently interact on Android. On the topic of analysis of permission-based architectures,

3. PROPOSED SOLUTION

As part of a solution to the above identified pitfalls in the android security model, we propose a reputation based security trust model to evaluate and validate the applications prior to installation. We have also analysed the consequences of a malicious application that has managed to get installed with the full consent of the end user. The Internet is full of genuine and malicious applications. An Android mobile owner can download different applications with varying reputation ratings. In this model, it is proposed that after downloading and before installing, the mobile device asks the AM Cloud for the reputation of the downloaded application.

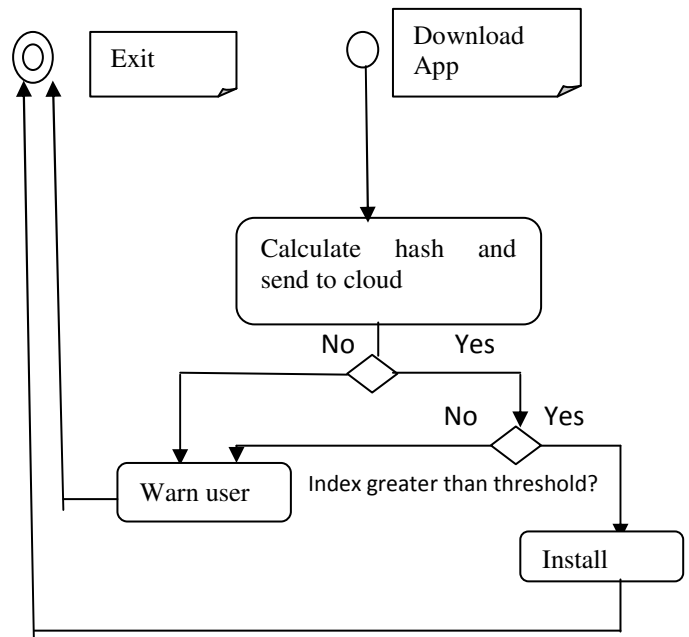


Fig. 1. overview of the proposed protocol

Based on the downloaded applications' behaviour and reputation index the downloaded application can be classified in any of the following three ways.

1. The application has built a good reputation and there is likely no harm installing it on the client's device. Good reputation will be set after some threshold of positive

feedback from those clients that have downloaded and automatically reported.

2. The application has not yet developed any good or bad reputation in the AM Cloud. In general, if an application has not developed a good reputation, we should be extremely cautious with such an unknown application. In this scenario, the anti-malware provider may wish to recommend that the user does not install the application or that the user installs the application in a sandbox.
3. The application has a bad reputation. In this case, the user is warned about the application's bad reputation.

4. EXPERIMENTS

Concerning just the applications which have not yet developed a strong reputation, we need to analyse those applications. To analyse the behaviour of an Android application, it is easier to start with analysing the set of permissions that the application has set in the Android application package file which includes all of the application's code, resources, assets, and manifest file. To do this, we have experimented with a reputation based security model for Android applications. A second experiment was also done to analyse how a malicious application could track a mobile owners' location and report it to a third party. The results were achieved using two experiments.

4.1. Experiment-1

One solution which has been used by anti-malware vendors is to perform analysis of the application, on the Android platform. However the Android is low on resources, such as performance, battery life and main memory. So it makes more sense to perform the analysis in the AM Cloud. To overcome these issues, another solution which has been used by anti-malware providers is to upload the entire application for analysis (for each user). For our solution, we will minimize the uploading of applications to the AM Cloud. I.e., we do not want two users, with the same exact application, to both upload the same application. Our approach to minimize the uploading of applications now follows.

4.2. Experiment-2

In this second experiment, we have developed two applications namely Location Tracker, The Location Tracker application has ACCESS_FINE_LOCATION, ACCESS_COARSE_LOCATION, and ACCESS_COARSE_LOCATION permissions in the user permission manifest file of the application. The manifest file declares which permissions the application must have in order to access

protected parts of the API and interact with other applications [18]. It also declares the permissions that others are required to have in order to interact with the application's components [18]. The Location Tracker application implements a location listener class that returns the latitude and longitude of the present location by consulting the Location Manager, which provides access to the system location services. We can use the latitude and longitude to locate the associated geographic place such as the street address, hotel, and zip codes.

5. FURTHER DISCUSSION

Designing a permission-based system is a challenging task because system designers must anticipate what usage will be given to the permissions defined in their system. The analysis in this paper has helped to identify developer usage patterns in a real-world dataset of top Android applications. Additionally, there is a constant struggle to make the system highly configurable under different use-cases while maintaining a low level of complexity. Understanding how the permission model is used in practice can help in making modifications to improve currently deployed permission systems. Furthermore, our analysis shows correlations between several of the infrequently used permissions. We note that having finer-grained permissions in a permission-based system enables users to have detailed control over what actions are allowed to take place. Whether it is beneficial to provide finer granularity will depend on many factors within a particular environment, as it increases complexity and thus may have, for example, usability impacts on designers and end-users. In the case of Android, having 'too many' permissions impacts both developers and end users. Developers must understand which permissions are needed to perform certain actions; determining this is often non-trivial, even for 'experts'. While some enthusiastic developers might take the time to learn what each of the 110 or more permissions do and request them appropriately when needed, other developers might choose to simply over-request functionality to make sure their application works.

5.1 Possible Enhancements to Android

The Android permission model does not currently make use of the implied hierarchy in its namespace. For example, `a.p.SEND_SMS` and `a.p.WRITE_SMS` are two independent permission labels, instead of being grouped, for instance, under `a.p.SMS`. Android includes an optional logical permission grouping [9] that is used for displaying permissions with more understandable names (e.g., one of the groupings reads "Services that cost you money" instead of `a.p.CALL_PHONE`). This grouping, however, does not allow developers to hierarchically define permissions, which could potentially extend current Android-defined permissions to express more detailed functionality. In the case of Android particularly, a permission hierarchy would

allow for an extensible naming convention and help developers more accurately select (only the) needed features. One example would be a free application that displays ads from domains belonging to Admob. Currently a developer would include the ad code snippet, and request the `a.p.INTERNET` permission. This permission allows the application to communicate over any network and retrieve any data from any server in the world. A more fine grained hierarchical permission scheme could enable the developer to request the `a.p.INTERNET`. `ADVERTISING` (`.admob.com`) permission which could limit network connectivity to only download ads in static HTML from sub domains of Admob. A hierarchical permission scheme could help users understand why an application is requesting specific permissions, but more importantly, could help developer's better use the principle of least privilege. This modification is not backwards compatible with the currently deployed Android OS, therefore it might be better suited for an entirely new model instead.

5.2 Applicability to Other Permission-Based Systems

The methodology presented in this work has allowed us to understand how developers use the permission-based security model in Android. We believe that our methodology is applicable to explore usage trends in other permission based systems. A base requirement for the methodology to work is being able to display applications and associated permissions for this representation to be possible, the set of permissions requested by an application must be accessible. In the case of Android, the set is statically readable in a manifest, but other systems might have different implementations. Google's Chrome OS extension system [4, 10] uses an Android-like manifest and permissions to access advanced functionality, which makes this system a prime candidate for applying our methodology. An empirical study of a large set of third-party extensions using our SOM-based methodology could help identify what correlations, if any, are present in requesting permissions to open tabs, read bookmarks, etc. This may also be of use in addressing other security concerns raised in recent work [10].

6. CONCLUSION

We have introduced a methodology to the security community for the empirical analysis of permission-based security models. In particular, we analysed the Android permission model to investigate how it is used in practice and to determine its strengths and weaknesses. The Self-Organizing Map (SOM) algorithm is employed, which allows for a 2-dimensional visualization of highly dimensional data. SOM also supports component planes analysis which can reveal interesting usage patterns. We have analysed the use of Android permissions in a real-world dataset of 1,100 applications, focusing on the top 50

application from 22 categories in the Android market. The results show that a small subset of the permissions is used very frequently where large subsets of permissions were used by very few applications. We suggest that the frequently used permissions, specifically `a.p.INTERNET`, do not provide sufficient expressiveness and hence may benefit from being divided into sub-categories, perhaps in a hierarchical manner. Conversely, infrequent permissions such as the self-defined and the complementary permissions (e.g., `install/ uninstall`) could be collapsed into a general category. Providing finer granularity for frequent permissions and combining the infrequent permissions can enhance the expressiveness of the permission model without increasing the complexity (i.e., maintaining a constant over all permission count) as a result of the additional permissions. We hope that our SOM-based methodology, including visualization, is of use to others exploring independent permission-based models.

REFERENCES

- [1] A. Shabtai, Y. Fledel, U. Kanonov, Y. Elovici, S. Dolev, and C. Glezer. Google Android: A Comprehensive Security Assessment. In *IEEE Security & Privacy*, Volume 8, Issue 2, pp. 35–44, March–April 2010.
- [2] T. Bläsing, L. Batyuk, A.-D. Schmidt, S.A. Camtepe and S. Albayrak. An Android Application Sandbox system for suspicious software detection. In *Proceedings of 5th International Conference on Malicious and Unwanted Software (MALWARE 2010)*, Nancy, France, Oct. 19–20, 2010.
- [3] M. Ongtang, S. McLaughlin, W. Enck, and P. McDaniel. Semantically Rich Application-Centric Security in Android. In *Proceedings of the Annual Computer Security Applications Conference (ACSAC '09)*, Austin, TX, USA, December 6–10, 2009.
- [4] W. Shin, S. Kiyomoto, K. Fukushima, and T. Tanaka. Towards Formal Analysis of the Permission-Based Security Model for Android. In *Proceedings of Fifth International Conference on Wireless and Mobile Communications (ICWMC '09)*, Cannes/La Bocca, France, August 23–29, 2009.
- [5] P. Teufl, C. Orthacker, S. Kraxberger, G. Lackner, M. Gissing, A. Marsalek, J. Leibetseder and O. Prevenhieber. Android Market Analysis with Activation Patterns, In *Proceedings of 3rd International ICST Conference on Security and Privacy in Mobile Information and Communication Systems (MOBISec 2011)*, Aalborg, Denmark, May 17–19, 2011.
- [6] C. Orthacker, P. Teufl, S. Kraxberger, G. Lackner, M. Gissing, A. Marsalek, J. Leibetseder, and O. Prevenhieber. Android Security Permissions - Can we trust them? In *Proceedings of 3rd International ICST Conference on Security and Privacy in Mobile Information and Communication Systems (MOBISec 2011)*, Aalborg, Denmark, May 17–19, 2011.
- [7] J. Burns. Developing Secure Mobile Applications for Android—An Introduction to Making Secure Android Applications, http://www.isecpartners.com/files/iSEC_Securing_Android_Apps.pdf, Accessed on May 8, 2012.

-
- [8] E. Chin, A. Porter Feltn, K. Greenwood, and D. Wagner. Analysing the Inter-application Communication in Android, University of California, Berkeley, Berkeley, CA, USA.
 - [9] T. Vidas, D. Votipka, and N. Christin. All Your Droid Are Belong To Us: A Survey of Current Android Attacks, INI/CyLab, Carnegie Mellon University.
 - [10] Android Market, <http://www.android.com/market>, Accessed on May 13, 2012.
 - [11] Android permissions, <http://android.git.kernel.org/?p=platform/frameworks/base.git;a=blob;f=core/res/AndroidManifest.xml>. Accessed on May 13, 2012.
 - [12] A. Shabtai, Y. Fledel, and Y. Elovici. Securing Android-powered mobile devices using SE Linux. In *IEEE Security & Privacy*, Volume 8, Issue 3, pp. 36–44, May–June 2010.
 - [13] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani. Crow droid. Behaviour-Based Malware Detection System for Android. In *Proceedings of the Workshop on Security and Privacy in Smartphone's and Mobile Devices (SPSM'11)*, Chicago, IL, USA, October 17, 2011.
 - [14] L. Yihe. An Information Security Model Based on Reputation and Integrality of P2P Network. In *Proceedings of 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing*, Wuhan, Hubei, China, April 25–26, 2009.
 - [15] L. Qi. Network Security Analysis Based on Reputation Evaluation. In *Proceedings of 2011 International Conference on Information Technology, Computer Engineering and Management Sciences (ICM 2011)*, Nanjing, China, September 24–25, 2011.
 - [16] <http://developer.android.com/reference/android/content/Context.html>
 - [17] <http://developer.android.com/reference/android/content/Context.html>
 - [18] <http://developer.android.com/guide/topics/manifest/manifest-intro.html>, Accessed on May 13, 2012.
 - [19] H. Bing. Analysis and Research of Systems Security Based on Android, In *Proceedings of 2012 Fifth International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Zhangjiajie, Hunan, and January 12–14, 2012.
 - [20] B. Berger, M. Bunke, and K. Sohr, An Android Security Case Study with Bauhaus, in the proceedings of 2011 18th Working Conference on Reverse Engineering (WCRE), Limerick, October 17–20.

Computational Intelligence in Wireless Sensors Networks

Janani Rajaraman

*Electronics and Instrumentation Engineering
S C S V M V University, Kanchipuram
janurang@gmail.com*

Abstract: Wireless Sensor networks (WSN) is an emerging technology and have great potential to be employed in critical situations like battlefields and commercial applications such as building, traffic surveillance, habitat monitoring and smart homes and many more scenarios. One of the major challenges wireless sensor networks face today is security. While the deployment of sensor nodes in an unattended environment makes the networks vulnerable to a variety of potential attacks, the inherent power and memory limitations of sensor nodes makes conventional security solutions unfeasible. The sensing technology combined with processing power and wireless communication makes it profitable for being exploited in great quantity in future. In this paper, the architecture of WSN is provided as well as the Characteristics of WSN, wide variety of attacks in WSN and their classification. Paradigms of Computational Intelligence (CI) have been successfully used in recent years to address various challenges such as data aggregation, energy aware routing, task scheduling, security and localization. CI provides adaptive mechanisms that exhibit intelligent behaviour in complex and dynamic environments like WSNs.

Index Terms: Wireless Sensors Networks, Applications of WSN, Sensors Nodes. Computational Intelligence

1. INTRODUCTION

WSN are a new form of distributed computing which consists of large number of Sensor nodes, which are capable of Wireless Communication, data acquiring unit and transmitting them to the central point. In recent years, considerable advances in the field of WSN has resulted in the development of several applications, some of these have been successfully integrated in industrial applications, such as environmental monitoring, object and event detection, military surveillance and precision agriculture. Unlike cellular networks that deny service when too many phones are active in a small area, the vision of Mesh networking is based on strength of node density, and the interconnection of a WSN only grows stronger as nodes are added. As long as there is sufficient density, a single network of sensor nodes can grow to cover limitless area. WSN is different for traditional wireless networks such as Mobile Ad hoc Network, Cellular Network, Bluetooth, Wireless Local Area Network and etc. In these traditional networks, the tasks of organization, routing and mobility management are used to

optimize Quality of Service (QoS) and heighten bandwidth efficiency. These wireless networks aim to provide excellent throughput and delay characteristics under high mobile conditions.

2. ARCHITETURE OF WSN

Up to now plenty of researches in WSN have been conducted, and formed several main research platforms, the architecture of WSN commonly used by most of the research platforms as shown in Fig1. The sensor nodes randomly dispersed in the monitoring area by aircrafts spreading, manually deployed rockets ejecting, constitute a network through self organization method. Each of these nodes has the capability of collecting data and routes data back to the base stations and the base stations send the information to the centre through internet. The end users can browse and process data through internet from the center.

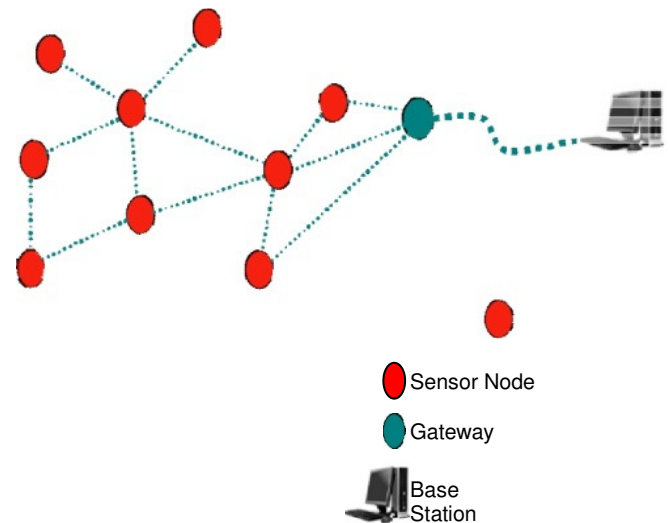


Fig. 1. Architecture of Wireless Sensor

3. COMPONENTS OF WSN

The main components of sensor network consist of a sensing unit, a processing unit, a transceiver, and a power unit as shown in Fig.2 [2]. A *transceiver* module is a class of device

that is not packaged within an enclosure. A transceiver module usually combines an embedded processor, radio stack, antenna and sometimes sensors and actuators, usually not packaged in an enclosure. Motes and the SHIMMER baseboard are two examples of transceiver modules. A *Sensor or Actuator Module* combines the sensing and I/O on a plug-in board for the transceiver module. By combining a transceiver module, a sensor module, an enclosure and a battery one would get a WSN device. A Sensor is the small piece of technology that actually interacts with the environment and which sends an appropriate signal to the embedded processor (Microcontroller Unit). Depending on the design and the choice of technologies, the sensor may be within the same transceiver module as the processor, or it may be plug in addition to the transceiver module. The microcontroller unit (MCU) may decide to forward the sensed signal to an aggregator, or to do some processing, sleep for a while, or wait until the next cycle to forward the information from the sensor. When ready, the MCU sends the signal to the radio stack; the radio stack then uses a communications protocol, and a transport protocol with a data protocol to pass the information to an aggregator. The data is sent out of the radio stack to an antenna and thereby received by another antenna. The communication protocol layers transmit on a given frequency and format.

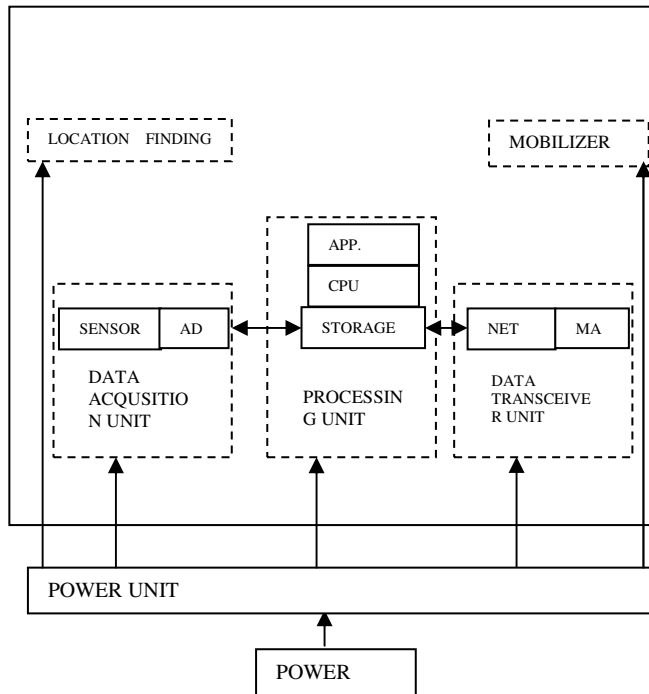


Fig 2. Components of a node in Wireless Sensor Networks

The *processing unit* plays a major role in managing collaboration with other sensors to achieve the predefined tasks. There are currently several families of this unit including microcontrollers, microprocessors and FPGA. Non

volatile memory interfaces such as ADCs can be integrated onto a single integrated circuit. The processing unit needs storage for tasking and to minimize the size of transmitted messages by local processing and data aggregation. *Power* consumption is a major weakness of sensor networks. Any energy preservation schemes can help to extend sensors life time. Batteries used in sensors can be categorized into two groups, rechargeable and non-rechargeable. Two major power saving policies can be found. Unused devices can shut down and activated when required. This is called “Dynamic Power Management”.

4. TYPES OF WSN

Terrestrial WSNs typically consist of hundreds to thousands of inexpensive wireless sensor nodes deployed in a given area, either in an ad hoc or in a pre-planned manner. In ad-hoc deployment, sensor nodes can be dropped from a plane and randomly placed into the target area. In pre-planned deployment, there is grid placement, optimal placement, 2-d and 3-d placement and models. Here energy can be conserved with multi-hop optimal routing, short transmission range, in-network data aggregation, eliminating data redundancy, minimizing delays, and using low duty cycle operations.

Underground WSNs and consist of a number of sensor nodes buried underground or in a cave or mine used to monitor underground conditions. Additional sink nodes are located above ground to relay information from the sensor nodes to the base station. The underground environment makes wireless communication a challenge due to signal losses and high levels of attenuation. Unlike terrestrial WSN, the deployment of an underground WSN requires careful planning and energy and cost considerations. Like terrestrial battery power and once deployed into the ground, it is difficult to recharge or replace a sensor node’s battery. A key objective is to conserve energy in order to increase the lifetime of network which can be achieved by implementing efficient communication protocol.

Underwater WSNs consist of a number of sensor nodes and vehicles deployed underwater. As opposite to terrestrial WSNs, underwater sensor nodes are more expensive and fewer sensor nodes are deployed. Autonomous underwater vehicles are used for exploration or gathering data from sensor nodes. Compared to a dense deployment of sensor nodes in a terrestrial WSN, a sparse deployment of sensor nodes is placed underwater. Typical underwater wireless communications are established through transmission of acoustic waves. Underwater sensor nodes must be able to self-configure and adapt to harsh ocean environment. The issue of energy conservation for underwater WSNs involves developing efficient underwater communication and networking techniques.

Multi-Media WSN have been proposed to enable monitoring and tracking of events in the form of multi-media such as video, audio, and imaging. Multi-media WSNs consists of a number of low cost sensors nodes equipped with cameras and microphones. These sensor nodes interconnect with each other over a wireless connection for data retrieval, process, correlation, and compression. Multi-media sensor nodes are deployed in a preplanned manner into the environment to guarantee coverage. Multi-media content such as a video stream requires high bandwidth in order for the content to be delivered. As a result, high data rate leads to high energy consumption.

Mobile WSNs consists of a collection of sensor nodes that can move on their own and interact with the physical environment. Mobile nodes have the ability sense, compute, and communicate like static nodes. A key difference is mobile nodes have the ability to reposition and organize itself in the network. A mobile WSN can start off with some initial deployment and nodes can be spread out to gather information. Information gathered by a mobile node can be communicated to another mobile node when they are within range of each other.

5. APPLICATIONS OF WSN

WSN may consists of many different types of sensors such as seismic, low sampling rate magnetic, thermal, visual, infrared, acoustic and radar, which are able to monitor a wide variety of ambient conditions that include temperature, humidity, vehicular movement, lightning condition, pressure, soil makeup, noise levels, the presence or absence of certain kinds of objects, mechanical stress levels on attached objects, and the current characteristics such as speed, direction, and size of an object. WSN can be used for continuous sensing, event detection, event IDm location sensing, and local control of actuators. Applications of WSN can be categorized into many fields such as military, environmental, healthy, home, commercial, and industrial areas.

A. Military Applications:

WSN can be an integral part of military Command Control, Communications, Computing, Intelligence, Surveillance, Reconnaissance and Targeting systems and is competent for monitoring friendly forces, equipment and ammunition, and other tasks. The rapid deployment, selforganization and fault diagnosis characteristics of WSN make it a very promising sensing technique for these military applications.

B. Industrial Applications:

Industrial applications include robot control and guidance in automatic manufacturing environments; industrial process control and automation; smart structure with sensor nodes

embedded inside; machine diagnosis; factory instrumentation; local control of actuators; instrumentation of semiconductor processing chambers; monitoring or rotating machine; monitoring of wind tunnels and anechoic chambers and etc.

C. Environmental Applications:

Environmental applications include tracking the movements of birds, small animals and insects; monitoring environmental conditions that affect crops and livestock irrigation; forest fire detection flood detection etc. [8]. In forest fire detection, since millions of sensor nodes can be strategically, randomly, and densely deployed in a forest, integrated using RF/Optical systems, sensor nodes can relay the exact origin of the fire to the end users before the fire is spread uncontrollable. [7]:

D. Healthy Applications

WSN can provide integrated patient monitoring diagnostics; drug administration in hospitals; monitoring the movements and internal process of infection; tele-monitoring of human physiological data and etc.[9]. If sensor nodes can be attached to medications, the probability of getting and prescribing wrong medications to patients will be minimized, as patients will have sensor nodes that identify their allergies and required medications.

E. Security and Surveillance

The important application of sensor networks is in security monitoring and surveillance for buildings, airports, subways, or other critical infrastructure such as power and telecom grids and nuclear power plants. Sensors may also be used to improve the safety of roads by providing warning of approaching cars at inter authenticate users. Imager or video sensors can be very useful in identifying and tracking moving entities, although they require higher bandwidth communication links. Heterogeneous systems that comprise both imagers and lower-cost sensors such as motion or acoustic sensors can be very cost effective; in these systems, lower-cost sensors can act as triggers for imagers. The security and reliability of systems themselves are essential, given the importance of critical infrastructure they are designed to protect.

6. CHALLENGES OF SECURITY IN WSN

The first challenges of security in sensor networks lie in the conflicting interest between minimizing resource consumption and maximizing security. Therefore the usefulness of a potential solution depends how good the compromise it achieves is. The *resource* in this context includes energy as well as computational resource like CPU cycles, memory, and communication bandwidth. Any

security mechanisms for WSN should take the following has five major resource constraints into consideration: (1) limited energy, (2) limited memory, (3) limited computing power, (4) limited communication bandwidth, (5) limited communication range; more or less in descending order of acuteness. Secondly, the capabilities and constraints of sensor node hardware will influence the type of security mechanisms that can be hosted on a sensor node platform. Since the amount of additional energy consumed for protecting each message is relatively small, the greatest consumer of energy in the security realm is key establishment. [3] Thirdly, the ad-hoc networking topology renders a WSN susceptible to link attacks ranging from passive eavesdropping to active interfering. Unlike fixed hardwired networks with physical defence at firewalls and gateways, attacks on a WSN can come from all directions and target at any node. Damage can include leaking secret information, interfering message and impersonating nodes, thus violating the above security goals. Fourthly, the wireless communication characteristics of WSN render traditional wired-based security schemes impractical.

Based on the above analysis on the security challenges, challenges and potential attacks in WSN, we further summarize three key issues for achieving the security of ad hoc networks:

(1) Key Management in WSN

Confidentiality, integrity, and authentication services are critical to preventing an adversary from compromising the security of a WSN. Key management is likewise critical to establishing the keys necessary to provide this protection in WSN. However, providing key management is difficult due to the ad hoc nature, intermittent connectivity, and resource limitations of the sensor network environment.

Traditional key management service is based on a trusted entity called a certificate authority (CA) to issue public key certificate of every node. The trusted CA is required to be online in many cases to support public key revocation and renewal. But it is dangerous to set up a key management service using a single CA in a sensor network. The single CA will be the vulnerable point of the network. If the CA is compromised, the security of the entire network is crashed. How to set up a trusted key management service for the WSN is a big issue.

(2) Securing routing of WSN

There are two kinds of threats to ad hoc routing protocols [03]: (1) External attackers. The attacks include injecting erroneous routing information, replaying old routing information, and distorting routing information. Using these ways, the attackers can successfully partition a network or introduce excessive traffic load into the network, therefore

cause retransmission and ineffective routing. Using cryptographic schemes, such as encryption and digital signature can defend against the external attacks. (2) Internal compromised nodes. They might send malicious routing information to other nodes. It is more severe because it is very difficult to detect such malicious information because compromised node can also generate valid signature.

Existing routing protocols cope well with the dynamic topology, but usually offer little or no security measures [8]. An extra challenge here is the implementation of the secured routing protocol in a network environment with dynamic topology, vulnerable nodes, limited computational abilities and strict power constraints.

(3) Prevention of Denial-of-service

Strictly speaking, although we usually use the term Denial-of-service (DoS) to refer to an adversary's attempt to disrupt, subvert, or destroy a network, a DoS attack is any event that diminishes or eliminates a network's capacity to perform its expected function. Hardware failures, software bugs, resource exhaustion, environmental conditions, or any complicated interaction between these factors can cause DoS.

An adversary may possess a broad range of DoS attack capabilities in WSN. For example, a wireless sensor network can be aerially deployed in enemy territory. If the enemy already has a wired network and power grid available and can interact with the newly deployed sensor network, it can apply powerful back-end resources to subvert or disrupt the new network.

7. ATTACKS ON SENSOR NETWORKS

Wireless Sensor networks are vulnerable to security attacks due to the broadcast nature of the transmission medium.[10]. Wireless sensor networks have an additional vulnerability because nodes are often placed in a hostile or dangerous environment where they are not physically protected. Fig.3 shows the attacks [11] classification on WSN.

A. Passive Attacks

The monitoring and listening of the communication channel by unauthorized attackers are known as passive attack. The Attacks against privacy is passive in nature.

1). Attacks against Privacy

The main privacy problem is not that sensor networks enable the collection of information. In fact much information from sensor networks could probably be collected through direct site surveillance. Rather, sensor networks intensify the privacy problem because they make large volumes of

information easily available through remote access. Hence, adversaries need not be physically present to maintain surveillance. They can gather information at low-risk in anonymous manner. Some of the most common attacks against sensor privacy are

- Monitor and Eavesdropping
- Traffic Analysis
- Comouflage Adversaries.

B. Active Attacks

The unauthorized attackers monitors, listens to and modifies the data stream in the communication channel are known as active attack. The following attacks are active in nature.

1. Routing Attacks in Sensor Networks
2. Denial of Service Attacks
3. Node Subversion
4. Node Malfunction
5. Node Outage
6. Physical Attacks
7. Message Corruption
8. False Node
9. Node Replication Attacks
10. Passive Information Gathering



Fig. 3. Classification of Security Attacks on WSN

8. PARADIGMS OF COMPUTATIONAL INTELLIGENCE IN WSN

The Computational Intelligence (CI) is the study of adaptive mechanisms that enable or facilitate intelligent behavior in complex and changing environments. These mechanisms include paradigms that exhibit an ability to learn or adapt to new situations, to generalize, abstract, discover and associate. CI is also defined as the computational models and tools of Intelligence capable of inputting raw numerical sensory data directly, processing them by exploiting the representational parallelism and pipelining the problem, generating reliable and timely responses and withstanding high fault tolerance. Paradigms of CI are designed to model the aspects of biological intelligence. CI encompasses paradigms such as neural networks, reinforcement learning, swarm intelligence, evolutionary algorithms, fuzzy logic and artificial immune systems. These paradigms are briefly introduced in the following subsections

1) Neural Networks

The human brain, which possesses an extraordinary ability to learn, memorize and generalize, is a dense network of over 10 billion neurons, each connected on average to about 10,000 other neurons. Each neuron receives signals through synapses, which control the effects of the signals on the neuron. These synaptic connections play an important role in the behaviour of the brain. A NN consists of a network of neurons organized in input, hidden and output layers. In feed forward NNs, the outputs of a layer are connected as the inputs to the next layer [12]. NNs learn the facts represented by patterns and determine their inter-relationships. Learning is the process in which the weights of a NN are updated in order to discover patterns or features in the input data. Learning methods are classified into the following types: i) supervised learning, ii) unsupervised learning and iii) reinforcement learning. In supervised learning, a teacher presents an input pattern and the corresponding target output. Network weights are adapted in such a way that the error is minimized. The objective of unsupervised learning is to discover patterns in the input data with no help from a teacher. In reinforcement learning, the learning agent communicates with its own environment, but not with a teacher. NNs have been found successful in a wide range of applications such as power system stabilization, pattern classification, speech recognition, robotics, prediction and image processing.

2) Fuzzy logic

Classical set theory allows elements to be either included in a set or not. This is in contrast with human reasoning, which includes a measure of imprecision or uncertainty, which is marked by the use of linguistic variables such as *most*, *many*, *frequently*, *seldom* etc. This approximate reasoning is modelled by fuzzy logic, which is a multi valued logic that

allows intermediate values to be defined between conventional threshold values. Fuzzy systems allow the use of fuzzy sets to draw conclusions and to make decisions. Fuzzy sets differ from classical sets in that they allow an object to be a partial member of a set. For example, a person may be a member of the set *tall* to a degree of 0.8[12]. In fuzzy systems, the dynamic behaviour of a system is characterized by a set of linguistic fuzzy rules based on the knowledge of a human expert. Fuzzy rules are of the general form: *if* antecedent(s) *then* consequent(s), where antecedents and consequents are propositions containing linguistic variables. Antecedents of a fuzzy rule form a combination of fuzzy sets through the use of logic operations. Thus, fuzzy sets and fuzzy rules together form the knowledge base of a rule-based inference system. Antecedents and consequents of a fuzzy rule form fuzzy *input space* and fuzzy *output space* respectively, which are defined by combinations of fuzzy sets. Non-fuzzy inputs are mapped to their fuzzy representation in the process called fuzzification. This involves application of membership functions such as triangular, trapezoidal, Gaussian etc. The inference process maps fuzzified inputs to the rule base to produce a fuzzy output. A consequent of the rule and its membership to the output sets are determined here. The defuzzification process converts the output of a fuzzy rule into a crisp, non-fuzzy form. Popular inference methods that determine an approximate non-fuzzy scalar value to represent the action to be taken include max-min method, averaging method; root sum squared method, and clipped center of gravity method. Fuzzy logic has been applied successfully in control systems (e.g., control of vehicle subsystem, power systems, home appliances, elevators etc.), digital image processing and pattern recognition.

3) Evolutionary Algorithms

Evolutionary algorithms model the natural evolution, which is the process of adaptation with the aim of improving survival capabilities through processes such as natural selection, survival-of-the-fittest, reproduction, mutation, competition and symbiosis. EC encompasses a variety of EAs that share a common underlying idea of survival-of-the-fittest. EAs use a population of solution candidates called chromosomes. Chromosomes are composed of genes, which represent a distinct characteristic. A fitness function, which the EA seeks to maximize over the generations, quantifies the fitness of an individual chromosome. Process of reproduction is used to mix characteristics of two or more chromosomes (called parents) into the new ones (called offspring). Offspring chromosomes are mutated through small, random genetic changes in order to increase diversity. Some fittest chromosomes are selected to go into the next generation, and the rest are eliminated. The process is repeated generation after generation until either a fit-enough solution is found or a previously set computational limit is reached. Following are the major classes of EAs.

- Genetic algorithms, which model genetic evolution
- Genetic programming whose individual chromosomes are computer programs
- Evolutionary programming which model adaptive behaviour in evolution
- Evolutionary strategies which model strategy parameters that control variation in evolution
- Differential evolution which is identical to GA except for the reproduction mechanism
- Cultural evolution which models the evolution of culture of a population and culture's influence on genetic and adaptive evolution of individuals
- Co evolution in which initially "dumb" individuals evolve through cooperation or competition and become fit enough to survive Successful applications of EA include planning, design, control, classification and clustering, time series modelling, music composing etc.

4. Swarm Intelligence

SI originated from the study of collective behaviour of societies of biological species such as flocks of birds, shoals of fish and colonies of ants. SI is the property of a system whereby collective behaviours of unsophisticated agents interacting locally with their environment cause coherent functional global patterns to emerge. While graceful but unpredictable bird-flock choreography inspired the development of particle swarm optimization, impressive ability of a colony of ants to find shortest path to their food inspired the development of ant colony optimization. The honey bee algorithm mimics foraging behaviour of swarms of honey bees

1) Particle Swarm Optimization:

The basic PSO consists of a population (or a swarm) of s particles, each of which represents a candidate solution. The particles explore an n -dimensional space in search of the global solution, where n represents the number of parameters to be optimized. Each particle i occupies position x_{id} and moves with a velocity vid , $1 \leq i \leq s$ and $1 \leq d \leq n$. Particles are initially assigned random positions and velocities within fixed boundaries, i.e., $x_{min} \leq x_{id} \leq x_{max}$ and $v_{min} \leq vid \leq v_{max}$ (in most cases $v_{min} = -v_{max}$). Fitness of a particle is determined from its position. The fitness is defined in such a way that a particle closer to the solution has higher fitness value than a particle that is far away. In each iteration, velocities and positions of all particles are updated to persuade them to achieve better fitness. The process of updating is repeated iteratively either until a particle reaches the global solution within permissible tolerance limits, or until a sufficiently large number of iterations are reached. Magnitude and direction of movement of a particle is

influenced by its previous velocity, its experience and the knowledge it acquires from the swarm through social interaction.

2) Ant Colony Optimization:

ACO was introduced in as a meta heuristic for solving combinatorial optimization problems. Foraging ants initially explore surroundings of their nest in a random manner. When an ant finds a source of food, it evaluates quantity and quality of the food and carries some food to the nest. While returning to the nest, the ant deposits a trail of chemical pheromone, which guides other ants to the food source [13]. This characteristic of ant colonies is exploited in artificial ant colonies to solve combinatorial optimization problems. The main idea of the ACO meta heuristic is to model the problem as a search for the best path in a "construction graph" that represents the states of the problem. Artificial ants walk through this graph, looking for good paths. They communicate by laying pheromone trails on edges of the graph, and they choose their path with respect to probabilities that depend on the amount of pheromone previously left.

5. Reinforcement Learning

Conventional artificial intelligence is based on machine learning, which is the development of the techniques and algorithms that allow machines to simulate learning. Machine learning attempts to use computer programs to generate patterns or rules from large data sets. RL is a sub-area of machine learning concerned with how an agent should take actions in an environment so as to maximize some notion of a long-term reward. RL is biologically inspired and acquires its knowledge by actively exploring its environment. At each step, it selects some possible action and receives a reward from the environment for this specific action. Note that the *best* possible action at some state is never known a-priori. Consequently, the agent has to try many different actions and sequences of actions and learns from its experiences. Usually, reinforcement learning tasks are described as a Markov decision process, consisting of an agent, set of possible states S , set of possible actions $A(st)$ for all possible states st and a reward function $R(st, at)$, specifying the environment reward to the agent's selected action. Additionally, the *policy* π defines how the learning agent behaves at some time-step t . The optimal policy is usually defined as π^* . The *value function* $V(st, at)$ defines the expected total reward when taking action at in state st , if from the next state $st+1$ the optimal policy π^* is followed. This is the function the agent has to learn in order to achieve the optimal policy. RL is well suited for distributed problems, like routing. It has medium requirements for memory and computation at the individual nodes. This arises from the need of keeping many different possible actions and their values. It needs some time to converge, but is easy

to implement, highly flexible to topology changes and achieves optimal results.

9. CONCLUSION

Recent literature shows that researchers have focused their attention on innovative use of CI techniques to address WSN issues such as design and deployment, localization, security, optimal routing and clustering, scheduling, data aggregation and fusion, and QoS management. Recent implementations of CI methods in various dynamical and heterogeneous networks are presented in this survey paper. CI paradigms and numerous challenges in sensor networks are briefly introduced, and the CI approaches used by researchers to address the challenges are briefly explained. In addition, a general evaluation of CI algorithms is presented, which will serve as a guide for using CI algorithms for WSNs. An advanced CI approach called adaptive critic design holds promise to generate practical optimal/sub-optimal solutions to the distributed sensor scheduling problem. There are successful applications of adaptive critic designs in power systems, which show that the technique provides guaranteed stable optimal solutions under uncertainties and noise. The potential of adaptive critic designs remains to be exploited in the field of WSN scheduling. In spite of a multitude of successful CI applications in WSNs, the main concern is that the most of these algorithms or protocols are still in development stage, and they may remain forever in non-finalized state. Very few protocols have grown out of the simulation environment. Most of them do not even consider unreliable or asymmetric links, node failure and mobility. Besides, a common problem is the lack of comparison to conventional state-of-the-art protocols to clearly identify the advantages of introducing CI. Thus, the goal of CI research community for the future of CI in WSNs is to improve already existing solutions, refine them and define well-performing real-world protocols. There are only a few already published initiatives in this direction.

REFERENCES

- [1] I.F. Akyildiz, W.Su, Y. Sankarasubramaniam, and E.Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, 2002.
- [2] Ruizhong Lin, Zhi Wang, and Youxian Sun, "Wireless Sensors Networks Solutions for Real Time Monitoring of Nuclear Power Plant," *Proceedings on Intelligent Control and Automation*, June 2004.
- [3] Zheng Yan, (Networking Laboratory, Helsinki University of Technology), "Security in Ad Hoc Networks," available from <http://citeseer.nj.nec.com/536945.html>.
- [4] Carman, D., Kruus, P., Matt, B., "Constraints and Approaches for Distributed Sensor Network Security", NAI Labs T.R. #00-010, June 1, 2000.
- [5] L. Zhou and Z. J. Haas, "Securing Ad Hoc Networks", *IEEE Networks*, Volume 13, Issue 6 1999.

-
- [6] Fei Hu, Jim Ziobro, Jason Tillett, Neeraj K. Sharma., “Secure Wireless Sensor Networks: Problems and Solutions”
 - [7] David Culler, Deborah Estrin, Mani Srivastava, “Overview of Sensor Networks”, *IEEE Computer Society*, August 2004.
 - [8] Alberto Cerpa, Jeremy Elson, Deborah Estrin and Jerry Zhao, “Habitat Monitoring: Application driver for Wireless Communications technology”, *2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean, Costa Rica*, April 2001.
 - [9] J. M. Kahn, R. H. Katz, and K. S. J. Pister. “Next Century Challenges: Mobile networking for smart dust”. In *Proceedings of 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, PP 483-492, August 1999.
 - [10] Dr. G. Padmavathi, and D. Shanmugapriya, “A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks”, *International Journal of Computer Science and Information Security*, Vol. 4, No. 1 & 2, 2009.
 - [11] Undercoffer, J., Avancha, S., Joshi, A. and Pinkston, J. “Security for sensor networks”. In *Proceedings of the CADIP Research Symposium*, University of Maryland, Baltimore County, USA, year 2002
 - [12] Andries P. Engelbrecht, “Computational Intelligence An Introduction”, Second Edition.
 - [13] Dervis Karaboga, Selcuk Okdem, Celal Ozturk, “Cluster based wireless sensor network routing using artificial bee colony algorithm”, *Wireless Networks* (2012).

Wireless Telecommunication Technologies Vulnerability and Attacks on Wireless Systems

Ankur Agrawal¹, Utkarsh Sharma²

¹*Electronics and Communication Engineering
Jaypee Institute of Information Technology, Noida, India
ankur_family123@yahoo.com*

²*Electrical and Electronics Engineering NIT, Uttarakhand, Srinagar, India
uttkarsh7@gmail.com*

Abstract: Wireless communications are being driven by the need for providing network access to mobile or nomadic computing devices. The need for wireless access to a network is evident in current work environments. The wireless technologies are an active area of research and applications. In this paper, different wireless technologies and the vulnerabilities for Bluetooth, Wi Fi etc are discussed. They have become the focus of great attention by both businessmen and consumer users. we have also discusses in detail some of the attacks including highway war driving, breaking of WEP 128 key, wireless auto configuration algorithm.. and many more. Coming on to the topic, we also discuss the remedies to tackle with wireless attacks and the introduction deals with several questions relates to sensor networks. Thus, we provide a framework for realistic security analysis in wireless sensor networks.

IndexTerms: Sensor networks, sensor network aims, highway war driving, WEP 128 key, wireless auto configuration theorem, vulnerabilities and attacks, remedies.

1. INTRODUCTION

Sensor networks provide unique opportunities of interaction between computer systems and their environment. Their deployment can be described at high level as follows:

1) The *sensor nodes* measure environmental characteristics which are then *processed* in order to *detect events*.

2) Upon event detection, *some actions* are triggered. This very general description applies to *extremely security-critical military applications* as well as to such benign ones (in terms of security needs)as *habitat monitoring*. Considering the *Internet* as an example, it is *extremely* difficult to add security to systems which were originally designed without security in mind. The goal of security is to “protect right things in a right way”^[1]. Thus, careful analysis is needed concerning which things to protect against which threats and how to protect them. Of course, this analysis is only possible in context of a particular class of applications. However, it makes much sense to provide a set of abstract security requirements and a set of generic attacker models,

i.e., a *framework for security analysis in wireless sensor networks*, which can be refined for particular applications.

In this paper, we present such a framework, which provides concepts to clarify *two important aspects* of the security analysis in wireless sensor networks:

1. What should be protected?

Here we offer a set of generic classes of requirements which can be used to structure and refine a set of concrete security.

We highlight the main differences between security requirements in classical systems and security requirements in wireless sensor networks.

2. Against what are we protecting the system?

Here we offer a set of generic attacker models which can be used to choose and refine particular attacker models for individual systems.

3) Overall, attacker models in conjunction with security requirements determine the means to achieve security. In practice it is very important to formulate realistic security requirements and realistic attacker models. Such choices guarantee that precious resources of *wireless sensor networks* are invested efficiently. It is therefore useful to evaluate the practicality of certain attacker models. One metric to measure practicality is to evaluate the effort an attacker has to invest to *perform certain attacks*. We contribute to this area by reporting on a number of experiments in which we *attacked real sensor node* hardware. This chapter is structured as follows:

- We first give an overview of security Goals in sensor networks, i.e., we approach the question “what to protect”.
- We then report on experiments in attacking wireless sensor networks .

Building on these experiences we develop a generic set of attacker models i.e., we approach the question “against whom to protect”. Finally, we briefly discuss protection mechanisms, i.e., we approach the question “*how to protect*”. *IN THIS PAPER*, we outline the open problems and summarize

2. SENSOR NETWORK AIMS: “FOR SECURITY PURPOSES”

As already mentioned it is (sensor network) is a high profile distributed system.

THE MAIN GOALS OF IT INCLUDES:-

i) The data should be accessible only to authorized users (Confidentiality), the data should be genuine.

ii) The data should be always available on the request of an authorized user (Availability) the above are some major aims/goals provided by the *incredible network*, some other goals include:-

- 1) Query Processing
- 2) Access Control
- 3) Large Scale Anti Jamming Services

3. THE FORMER SECURITY ISSUES ARE DESCRIBED AS

Outside Security Whereas, the Latter Security Issues are Described as Inside Security.

These aims ensure realizes robust, confidential and authenticated communication between individual nodes. This also includes in-network processing, data aggregation, routing and in-network data storage.

To summarize, security goals in sensor networks are similar to security goals in distributed databases (outside security) and distributed systems (inside security). So these can be taken as an orientation. While requirements are similar, many standard mechanisms to implement security (e.g., public key infrastructures or agreement protocols) are not applicable because they require too many resources or do not scale to hundreds or thousands of nodes.

1. Attacking Wireless Sensor Networks

Of course, the many possibilities to attack WSNs include all the classical techniques known from classical system security.

Some of the popular attacks are :

Type of Attack	Description	Methods and Tools
War Driving	Discovering wireless LANs by listening to beacons or sending probe requests, thereby providing launch point for further attacks.	Airmon-ng, DStumbler, KisMAC, MacStumbler, NetStumbler, Wellenreiter, WiFiFoFum
Rogue Access Points	Installing an unsecured AP inside firewall, creating open backdoor into trusted network	Any hardware or software AP
Ad Hoc Associations	Connecting directly to an unsecured station to circumvent AP security or to attack station.	Any wireless card or USB adapter
MAC Spoofing	Reconfiguring an attacker's MAC address to pose as an authorized AP or station	MacChanger, SirMACsAlot, SMAC, Wellenreiter, wicontrol
802.1X RADIUS Cracking Recovering	RADIUS secret by brute force from 802.1X access request, for use by evil twin AP.	Packet capture tool on LAN or network path between AP and RADIUS server

4. CONFIDENTIALITY ATTACKS

These attacks attempt to intercept private information sent over wireless associations, whether sent in the clear or encrypted by 802.11 or higher layer protocols.

Type of Attack	Description	Methods and Tools
Eavesdropping	Capturing and decoding unprotected application traffic to obtain potentially sensitive information	bsd-airtools, Ettercap, Kismet, Wire shark, commercial analyzers
WEP Key Cracking	Capturing data to recover a WEP key using passive or active methods.	Aircrack-ng, airoway, AirSnort, chop chop, dwepercrack, WepAttack, WepDecrypt, WepLab, wesside
Evil Twin AP	Masquerading as an authorized AP by beaconing the WLAN's service set identifier (SSID) to lure users	cquireAP, D-Link G200, HermesAP, Rogue Squadron, WifiBSD
AP Phishing	Running a phony portal or Web server on an evil twin AP to "phish" for user logins, credit card numbers.	Airpwn, Airsnarf, Hot spotter, Karma, RGlueAP

5. WIRELESS AUTO CONFIGURATION ALGORITHM

- 1) Access Point within range responds with Probe Responses.
- 2) If Probe Responses are received for networks in preferred networks list connect to them in preferred network list order.
- 3) If no available networks match preferred networks specific probe request are sent for each network which are hidden.
- 4) If still not associated and there is an ad-hoc network in preferred list, create the network and become first node.
- 5) Finally, if "Automatically connect to non-referred networks" is enabled (disabled by default), connect to networks in order they were detected.
- 6) Otherwise, wait for user to select a network or preferred network to appear Set card's SSID to random 32-char value, Sleep for minute, and then restart algorithm.
- 7) Attacker spoofs dissociation to victim.
- 8) Client sends broad cast and specific Probe Requests again. Attacker discovers networks in Preferred Networks list (e.g. Inksys, MegaCorp, t-mobile)
- 9) Attacker creates a rogue access point with SSID

6. VULNERABILITY OF WIRELESS NETWORKS

6.1 Highway war driving

We conducted war-driving with laptop computers in some of the highways. We went to different areas where wireless networks were detected and started capturing packets using the pre-configured laptops we had. As in table 1, Net Stumbler 0.4.0 software (NSS, 2005) and Link Ferret 3.10 software (LFS, 2005) were used for network detection and packets capturing respectively. CISCO Aironet 350 series PCMCIA Wireless adapter was fixed and configured. WinPcap software was also needed to be installed for packet capturing to work. The Link Ferret software can be configured to capture packets from different channels with a huge buffer size, with average packet size of around 64 bytes or more tables

Table 1 List of war driving tools used

Equipment	Item Specification
Laptop	Acer Laptop with Mobile Centrino processor, 256 MB RAM and 20 GB HDD
2. Wireless Network Adapters	On board wireless network adapter and CISCO Aironet 350 series PCMCIA
3. Network Detection Software	NetStumbler 0.4.0

2.2) Breaking of WEP-128 key

802.11b test bed using Aircrack (version 2.1) software. WEP-128 key was cracked by capturing round 4 million packets containing 264674 unique IVs in 2 seconds as shown. Once the key is broken, various sensitive details can be known through decryption. That can include information to launch other DoS attacks too.

Some other attacks:

Availability Attacks:

Type of attack	Description	Method and Tools
Shared Key Guessing	Attempting 802.11 Shared Key Authentication with guessed, vendor default or cracked WEP keys.	WEP Cracking Tools
PSK Cracking	Recovering a WPA/WPA2 PSK from captured key handshake frames using a dictionary attack tool.	coWPAtty, genpmk, KisMAC, wpa_crack
Application Login Theft	Capturing user credentials (e.g., e-mail address and password) from clear text application protocols	Ace Password Sniffer, Dsniff, PHoss, Win Sniffer
Domain Login Cracking	Recovering user credentials (e.g., Windows login and password) by cracking NetBIOS password hashes, using a brute-force or dictionary attack tool.	John the Ripper, L0phtCrack, Cain
VPN Login Cracking	Recovering user credentials (e.g., PPTP password or IPsec Preshared Secret Key) by running brute-force attacks on VPN authentication protocols.	ike_scan and ike_crack (IPsec), anger and THC-pptp-bruter (PPTP)
802.1x Identity Theft	Capturing user identities from clear text 802.1X Identity Response packets.	Capture Tools
802.1X Password Guessing	Using a captured identity, repeatedly attempting 802.1X authentication to guess the user's password.	Password Dictionary
802.1X LEAP Cracking	Recovering user credentials from captured 802.1X Lightweight EAP (LEAP) packets using a dictionary attack tool to crack the NT password hash.	Anwrap, Asleep, THC-LEAPcracker
802.1X EAP Downgrade	Forcing an 802.1X server to offer a weaker type of authentication using forged EAP-Response/Nak packets.	File2air, libradiate

7. CONCLUSION

This paper explicitly shows different types of attacks, their descriptions, their methods with which they are implemented, the tools used and moreover the remedies which should be taken in order to avoid these attacks, this is typically due to the nature of wireless communication as packets frantically move around in the air.

8. REMEDIES

Download Prevent Wireless Attack Software in keywords

- 1) Air Stop Multiport Edition v.2.5.0communication adapter, control adapter, and prevent wireless attack, communication, adapter, wireless attack
- 2) The Wireless Toolkit wireless, Verizon wireless, wireless router, wireless phone, Cingular wireless, x

box wireless controller, wireless phones, wireless internet, wireless service, xbox wireless controller

- 3) PDF Protection Remover decrypt pdf, pdf password, prevent printing, prevent copying, prevent extracting, prevent changing, prevent editing, user password, owner password.

REFERENCES

- [1] Wade Trappe, Yan Yong Zhang, "Security Emerging Wireless Systems"
- [2] www.texasinstrumentss.com
- [3] www.nationalinstruments.com
- [4] www.electronicforum.com
- [5] http://en.wikipedia.org/wiki/Wireless_security
- [6] http://interscience.in/IJSSAN_Vol1Iss1/paper25.pdf

Software as a Service (SaaS) Approach for E-learning

M. K. Sharma¹, Manisha Gururani²

¹Associate Professor, ²Assistant Professor

¹Amrapali Institute, Haldwani (Uttarakhand), ²Amity University, Noida (U.P.)

¹sharmamkhld@gmail.com, ²mgururani@gmail.com

Abstract: The recent advancements in Information and Communication Technology (ICT) provides the better IT solutions to the real world problems. The education field has no exception. Emerging IT technologies and platforms are used frequently by the people in learning processes. Way of learning and education changing with the latest advancements in Information and Communication Technology. The latest computing paradigm, that is cloud computing can be the one of the best way of computing technology in higher education and e-learning. In this paper we describes a proposed architecture model of academic cloud and gives the idea how we can offer e-learning services using SaaS model of cloud computing as a service. This describes how the cloud computing can be used effectively in e-learning process and sharing of resources of the academic institutes and providing the close overlook on protection of open access networks in higher educational institutions and the vision of technology for education.

IndexTerms: e-learning, cloud computing, academic cloud, e-learning as a service, SaaS,

1. INTRODUCTION

The use of technology is gradually increasing in Higher education specially Information & Communication Technology. The printed books and materials are replaced by the digitized books, animations, video lectures which gives a better experience to the user. These technology leads to greater benefits as in cost reduction, enhanced learning outcome and also in environment protection by reducing the use of papers. Cloud computing [1] [2] is one of them which provides the service oriented computing platforms and greater flexibility in accessing and delivering resources over the network. As the cloud computing provides the way to access the content and deliver the content across the world, so it can be used in educational institutions for delivering the e-contents and sharing the resources such as digital libraries. This can be said as e-learning as service i.e. the user can access the contents and learning materials as per use.

2. BENEFITS OF E-LEARNING

A. Reduced Cost of Learning Materials- The cost reduction is the greater benefit of e-learning. The e-contents are cheaper than the printed e-books and can be made no of

copies and shared among the people with less effort by the use of academic cloud approach i.e. e-learning as service.

B. Increased Participation of academic institutes- as the e-contents can be shared, stored and manipulated over the network which provides the increased participation of academic institutes. The two or more no of institutes can be connected and their resources can be shared which will help to enhanced technology learning.

C. Greater accessibility and better learning outcome- The e-learning provides the greater accessibility among the users and better learning outcome. The concepts and ideas can be represented in more attractive and clear way by using animation and graphics.

D. Flexibility- As we are using the cloud computing in this approach which will provides the flexibility to select the courses, materials and e-learning services to the users.

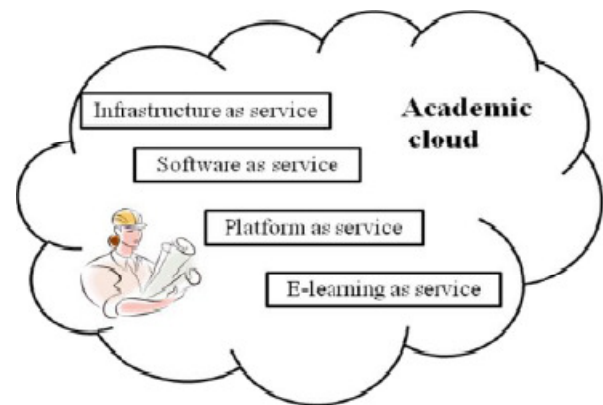


Fig. 1. Academic E-learning using cloud environment

3. ACADEMIC CLOUD

Academic cloud [3] is the use of cloud computing in academic institutions for enhancing the learning outcome, sharing of resources, and reduction in cost of maintenance. The cloud computing provides the three services i.e. Infrastructure as service, software as service, platform as

service and now it comes up e-learning as service for academic institutes.

Educational institutions throughout the World have become highly dependent on information technology to service their business requirements. Procuring and maintaining a wide range of hardware and software require substantial, ongoing investment and the skills to support them. The economies of scale and other features of cloud computing are likely to mean an increasing shift away from institutionally-hosted services. These services are increasingly provided using Internet technologies to staff and students and accessed from web browsers. The services are offered cheaply or freely to education, often with much higher availability than can be provided by the educational institution.[4][5]

4. PROPOSED MODEL

The proposed scheme focuses on the e-learning as service and how the resources of organizations can be shared and connected together to perform the all activities with greater flexibilities. Our proposed academic cloud architecture consists of cloud service providers such as Amazon EC2 [6], Nirvanix [7] and salesforce [8], local servers and central cloud system. According to our proposed architecture the each individual computers in the organization acts as the cloud partner and connected with the central cloud system by the local server. The local server provides the access of all computing resources and e-learning materials. In this the students, faculty members and other staffs can connect together and discuss, share their resources and ideas. The whole organizations can be connected such as academic block, library, laboratory, bank, hospital and hostels. Whenever we need to access the resources we can connect with the central system by local servers.

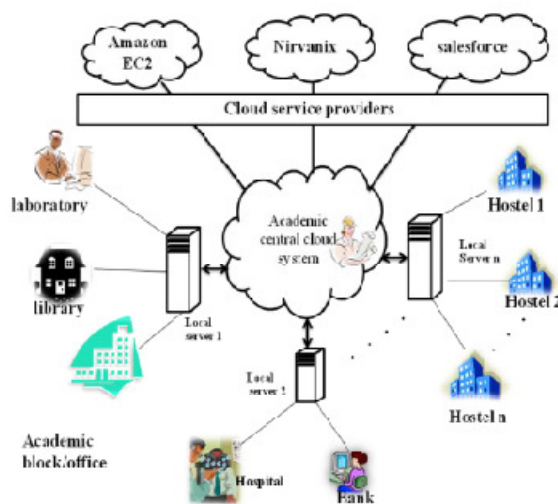


Fig. 2. Academic Cloud Architecture for an Institute

Communication between user and local server- The user and local server communicates together as request response method. The user sends the request for accessing the resource and based on the service provided the local server gives a acknowledgment. After receiving the acknowledgment the connection will be established between the user and local server and resource sharing takes place. For security purpose we can use the user ID and passwords so that only the valid users can access the provided resources by the organization. The security provisions can be provided and included in the system so that the access rights can be maintained properly. The students and staffs rights will differ such as a student can take the online exams, can send feedback, and can send home works and projects. In similar way the staff can prepare the test, can deal with content management, can assess tests, homework, projects taken by students and can send the feedback to their parents.

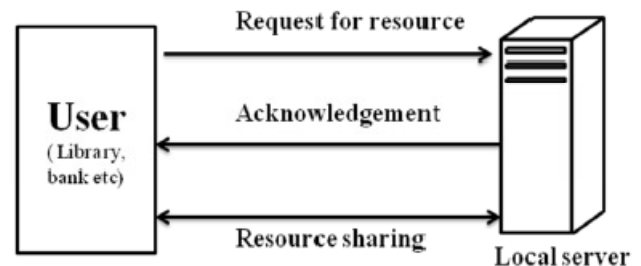


Fig. 3. communication process between user and local server

Steps of the procedure-

1. User (student, staff etc) request is sent to the local server with necessary user id and password.
2. Authentication module of local server checks and verifies the user login id and password and redirects to the next information page.
3. After receiving the specification from the local server verifies the current available resources like pricing, data security etc.
4. If the user does not have an agreement to receive the required services or if the pricing policy not matched then the local server immediately inform the user for payment by credit card or debit card and redirects to the payment information page.
5. If user is agree with the policy then the acknowledgement message sent to the local server.

5. CONNECTIVITY OF ACADEMIC INSTITUTIONS

The all the academic institutions can be connected globally and they can share the resources and e-contents for e-

learning process. [9] The figure below shows how the academic institutes can be connected for e-learning process by the academic cloud. Cloud provides the opportunity of flexibility and adaptability to use the computing resources on-demand. Contrary to having only one service provider, different providers use different interfaces to their compute resources utilizing varied architectures and implementation technologies for customers. Although this creates a management problem, a common architecture facilitates the management of compute resources from different Cloud providers in a homogenous manner (Dodda, Smith & van Moorsel, 2009). Mitchell (2008) provided an overview of existing learning architectures, and raised questions about how educational institutions are managing the cloud computing resources. He also brought reasonable explanations for the challenge of indexing web resources for optimum discoverability by students and educators.

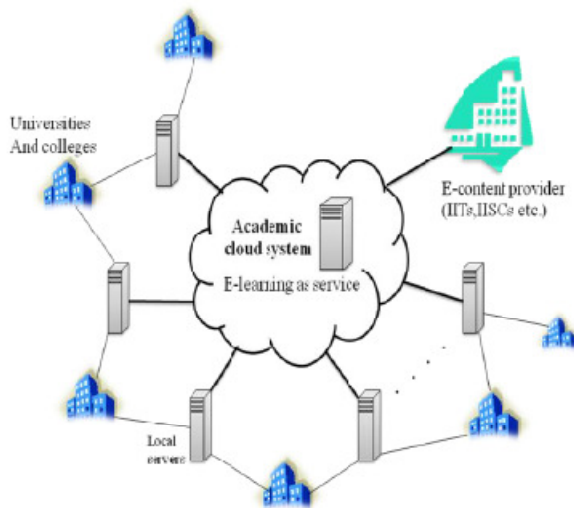


Fig. 4. connecting academic institutions

6. BENEFITS OF PROPOSED SCHEME OVER EXISTING METHODS

E-learning is widely used on different educational levels such as continuous education, company trainings, and academic courses. The proposed scheme includes the benefits such as:

i) Cost reduction – This method greatly reduces the cost of providing e-learning solutions compared to existing method. In our proposed architecture the cloud system has the knowledge of all unused resources and assigns resources from those free portions if any request arrives. In our designed architecture the partners associated with the cloud system offers software to the cloud for individual user with a reasonable price rate. The use of virtualization can enable the reduction in the hardware cost and energy consumption by the organizations computing infrastructure.

ii) Smarter Classroom- This proposed scheme can enable the classrooms to make smart and virtual. The all necessary resources can be integrated together and shared effectively. The staff can deliver their lectures from one place which will be made accessible by the students anywhere across the institution or globe through the global connectivity. The previous lectures can be stored in the database attached with the local server and can be made accessible to the students later when they need to access. The access rights will be maintained at each phase of resource sharing and content delivery.

iii) Data Portability- Data portability is very important in educational sectors. Student does several projects and research works on diverse fields. They collect several materials associated with their study like lecture, slides and various supplementary documents. This scheme enables to share all those slides and supplementary materials with their friends and faculty members. They can get the suggestions from them regarding their project works which is enabled only by this major advantage i.e. Data Portability.

iv) Smart Administration- This proposed method can enable the control of all resources effectively and in smart way since the entire system connected together and supported by the cloud service providers such as amazon EC2, Nirvanix, and Sales force.com. these providers will provide the all necessary support for teaching and other learning activities. The regular activity can be monitored properly which is done by the users and their suggestions and feedbacks can be collected. This will enable to know the user satisfaction and enables the enhancement in providing quality technical education by making the change in the system as per user's requirements.

v) Innovation in Research- Since this method contains the benefits of resource sharing and effective content management which will help to promote the research activities. The research needs the greater accessibility of the contents and resources without any breakage such as accessing the digital library and sending and receiving the suggestions and messages from the experts in the specific area of individual's interests. The global connectivity enables it happen.

7. CHALLENGES AND SECURITY ISSUES

Education is at a crossroads, which means our entire country is at a crossroads. Will we continue to spend tax dollars on older technologies that only perpetuate the problems? Will we continue to measure student success with methods that do little to improve learning in real time? Isn't it time to provide educators with real tools that improve learning in the classroom, rather than just trendy gadgets or yet another client-server silo?

Educators need connectivity. They need reliable, expansive communication tools and collaboration with the greater learning community, including parents, business leaders and civic leaders. With cloud computing, together we can innovate our way to success. Why not accept the challenge. [10]

The top four breaches of today

According to Paul Judge, chief research officer for Barracuda Networks[11], a security vendor based in Campbell, Calif., data security breaches in today's day and age generally fall into four basic categories.

Malicious Javascripts

This type of data security breach is by far the most prevalent, and comes in a variety of flavors. One, dubbed "malvertising," works through Web site ads. A hacker writes a program that he or she places behind an advertisement or other page on a perfectly legitimate Web site somewhere on the Internet. Users don't even have to click or hover over the ad to activate the malicious code; all they have to do is stop by. In many cases, these threats reside on perfectly trusted servers—servers that have been compromised at some time in the past.

Search engine malware

This strategy is a spin off malicious Javascripts—only instead of tricking users with a bad script on an otherwise harmless Web site, the approach feeds (the technical term is "poisons") search engines with malicious links. To get the biggest bang for their bucks, hackers stuff search engines with these links based on search topics that are trending (think "LeBron James," or "adjustable-height desks.") According to Judge, one in five search topics and one in 1,000 search results lead to malware of some kind.

Web Exploit Kits-

Finally, exploit kits are ready-made tools designed to help hackers attack vulnerable software (and sometimes hardware) components. Anyone (including run-of-the-mill college students) can buy them. Anyone (again, including students) can use them. Most kits come with dashboards that track the havoc a particular kit is wreaking. Perhaps most alarmingly, many kits are available on file-sharing sites—a kit named Black hole was found on a number of these sites earlier this spring.

Social attacks

In recent months, hackers also have exploited the social

nature of social networking sites such as Face book and Twitter. In the Face book environment, attacks are designed to look like photo tags, chats and seemingly innocuous apps (a recent one promised users two free tickets on Southwest airlines). Another common iteration: a phenomenon known as "Likejacking," whereby a hacker makes it look as if a trusted confidante has "liked" something that's actually a malicious script. In the Twitter environment, attacks appear in the form of fake links.

8. CONCLUSIONS AND FUTURE WORK

Cloud computing is an emerging computing paradigm which provides the variety of services in such a way that has not been experienced before. E-learning as a service can be one among them. The main aim of my this proposed architecture is to use the limited resources, which is available in a efficient and effective way to balance the current institutional resources in more economical way. The architecture proposed can be used by the developing country like India for improving their quality of higher education system. In future I will introduce a prototype of my proposed cloud based e-learning architecture and will discuss the practical security and implementation issues of this approach for educational Institutes. As there is no limitation in technological developments, the new advances will come in light in the area of cloud computing which will give a better user experience to next generation in e-learning process.

REFERENCES

- [1] Cloud Computing in Institutions, A briefing paper by Wilbert Kraan and Li Yuan.
- [2] <http://netseminar.stanford.edu/seminars/Cloud.pdf>
- [3] Education for a smarter planet: Cloud computing in Education by Alex caplen, IBM Cloud academy.
- [4] Case Study, Intel Education, Cloud computing brief, Schools, IT, and Cloud computing The Agility for 21st century e-learning.
- [5] Cloud computing for education: A new dawn? by Nabil Sultan* Faculty of Business and Computer Science, Liverpool Hope University, Hope Park, Liverpool, L16 9JD, UK, International Journal of Information Management.
- [6] <http://aws.amazon.com/ec2/>
- [7] <http://www.nirvanix.com/>
- [8] <http://www.salesforce.com/in/>
- [9] Demystifying cloud computing for higher education by Richard N.Katz, Phillip J. Goldstein, and Ronald Yanosky Research Bulletin ECAR.
- [10] The Promise of the Cloud for Education Understanding cloud computing and its true potential for educators by Tim Youngblood and Ken Mc Elrath.

Survey on Security Issues in Cloud Environment

Sudhir Shenai¹, M. Aramudhan², A. Devi Priya³

¹Department of CSE, Sathyabama University, Chennai, Tamil Nadu, shenai.sudhir@gmail.com

²Department of IT, PKIET, Karaikal, ponaranagai@yahoo.co.in

³AITTM, Amity University, Noida, cadevipriya@gmail.com

Abstract: Cloud Computing is a new wave in the Internet Revolution. It provides wide variety of services to users on rental, on-demand basis mostly by sharing the cloud vendor's physical infrastructure through virtualization. The main concern for wide adoption of cloud computing is security. Apart from the existing well-known security threats inherent to networking and computing environments, Cloud is vulnerable to some Cloud Specific Security (CSS) threats too. This paper analyses the various security threats common to cloud based services and general Internet services. The main focus is to explore Cloud Specific Security (CSS) vulnerabilities, threats and possible counter measures.

Index: Cloud Computing, CSS threats, CSS vulnerabilities

1. INTRODUCTION

First of all, just take a look how people concern about cloud computing. As you can see in this graph, people have been searching a lot for cloud computing in Google since 2007 especially since the first months of 2009. And other kind of computing such as grid computing is falling and cluster computing is still the same for years.

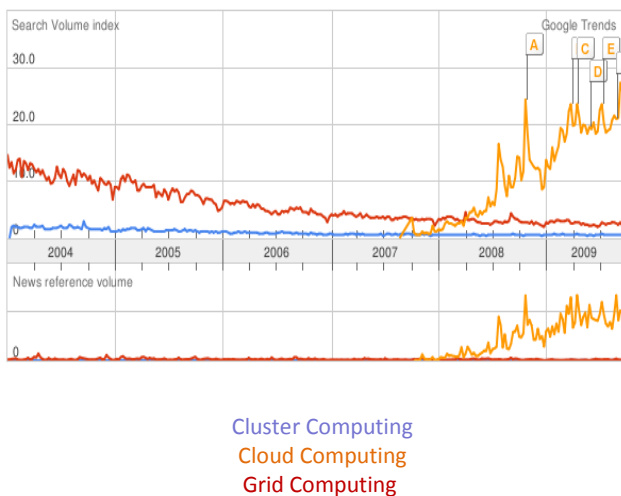


Fig.1 Search Volume Index of Cloud Computing-Google Trends

The graph shown above is the search volume index of 3 keywords cluster computing, grid computing and cloud computing in Google search trends from 2004 to 2009. This

hype cycle graph is a clear indicator of people's interest in cloud computing adoption. Along with the hype of cloud adoption, the fear of adoption was also growing due to the security vulnerabilities inherent to shared environment of cloud services. The major concern of cloud service providers is to convert this hype cycle to reality by wading away the security related apprehension from the potential customers, who will enable wide spread cloud adoption. Thus this paper is significant from the point of view cloud security researchers, cloud service providers and cloud service consumers. This paper analyses various security concerns related to cloud adoption, which are generic as well as cloud specific.

Let's start by demystifying the definition of Cloud Computing: There are many views for "what is Cloud Computing?" Over 20 Definitions can be found in http://cloudcomputing.sys-con.com/read/612375_p.htm. The Cloud Computing can be defined from various perspectives viz... Consumer Perspective, Service Provider Perspective and Business perspective etc...The definition by NIST covers a comprehensive view of all the perspectives. Hence we state here NIST's definition for Cloud Computing.

"Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources(e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"

The introduction chapter of this paper gives a brief idea about global interest in cloud computing and its definition in a nut shell. This remainder of the paper is structured as follows: Section 2 gives an overview of the survey by various organizations, which express the global interest in Cloud Adoption, its barriers and apprehensions. This gives an idea about area of security measures needed to be addressed for wide spread cloud adoption. Section 3 discusses the implications of common attacks thriving in other environment over cloud environment. Section 4 explores the cloud specific security threats, vulnerabilities and attacks. Section 5 explores the possible counter measures necessary for a secure cloud environment.

2. CLOUD SECURITY SURVEY

As per the 2012 Cloud Computing Security Survey [24] [25] by Information Security Media Group, nearly 1 in 3 survey respondents say their organizations are not using the cloud, a strikingly high percentage considering how quickly the computing platform is maturing. By a 72 percent-to-28 percent margin, the survey respondents say concerns about security prevent their organizations from adopting cloud services. The greatest reservations about secure Cloud Computing as per the survey are shown in the chart below.

An online survey was conducted by Trend Micro Inc. in June 2012 in seven countries: the US, Canada, the UK, Germany, Japan, India, and Brazil. Overall, 1,400 respondents (200 per country) were surveyed: all were purchase decision makers or key influencers for one or more of the following three solutions/services: Cloud Computing Services, Server Virtualization, and/or Virtual Desktop Infrastructure (VDI). Companies surveyed were limited to 500+ employees. The objective of the survey is to understand more about global and regional virtualization and cloud adoption rates and barriers to adoption as well as percent-to-28 percent margin, the survey respondents say concerns about security prevent their organizations from adopting cloud services. The greatest reservations about secure Cloud Computing as per the survey are shown in the chart below.

An online survey was conducted by Trend Micro Inc. in June 2012 in seven countries: the US, Canada, the UK, Germany, Japan, India, and Brazil. Overall, 1,400 respondents (200 per country) were surveyed: all were purchase decision makers or key influencers for one or more of the following three solutions/services: Cloud Computing Services, Server Virtualization, and/or Virtual Desktop Infrastructure (VDI). Companies surveyed were limited to 500+ employees. The objective of the survey is to understand more about global and regional virtualization and cloud adoption rates and barriers to adoption as well as to explore trends within the area of cloud technology and virtualization practices.

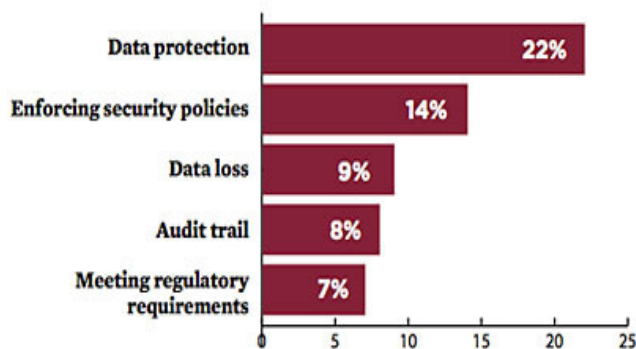


Fig. 2. 2012 Survey Results of Cloud Security by iSMG

Globally, 47% of the respondents who are currently using a cloud computing service reported they have experienced a data security lapse or issue with the cloud service their company is using within the last 12 months. India had the highest incidence (67%), followed by Brazil (55%).

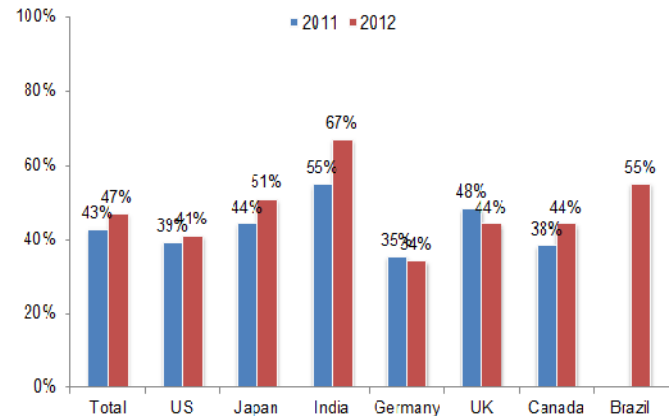


Fig. 3. Incidence of Data Security Lapse or issue with cloud service (%) Survey by Trend Micro Inc.-2012

With the adoption rate of cloud services increasing from 2011 to 2012, there was also an increase in the incidence of data security lapse or issue with the cloud services companies are using. This correlation is particularly strong in India and Canada.

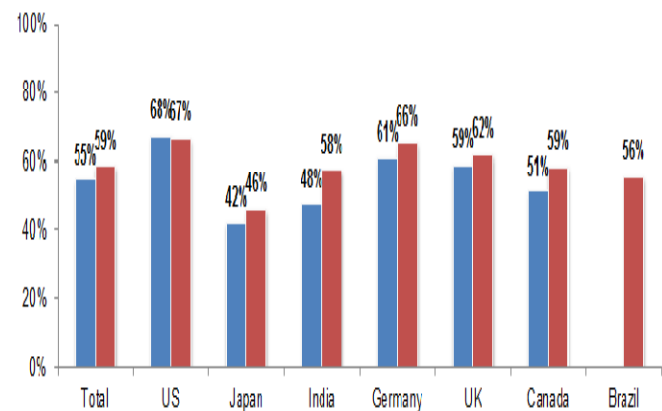


Fig. 4. Cloud Service Adoption Rate in (%) - Survey by Trend Micro Inc.-2012

The above survey gives a clear indication of high necessity of security lapses to be addressed for wide spread adoption of cloud services. It also gives a clear picture of how the security threats common to computing and networking environments raises its severity when it comes to cloud because of its shared nature and virtualization support. Also, it increases the loop holes for new variety of attacks which we term it as Cloud Specific Attack (CSA).

3. COMMON SECURITY ISSUES

An elaborate discussion of various security challenges related to cloud computing can be seen in the papers [4][5][6][7]. The security threats are classified based on the security requirements at various levels of services provided by the cloud [8]. Basically the cloud services are provided at Application Level, Platform Level and Infrastructure Level. The services provided at various levels differ from each other by virtue, hence the security requirements too. Further Cloud Security Alliance (CSA) classifies the security areas needed to be addressed in cloud computing [9]. Table I gives the summary of the classification of threats and security requirements at all the three levels of cloud services.

Many a kind of attacks are possible in the cloud such as the wrapping attack, Malware Injection Attack, Metadata Spoofing Attack [22], SQL injection attack, Cross site scripting, DDoS attack and DNS attack. But all the attacks are not specific to the Cloud. Most of the attacks are

common to other environments too but the virility of the same attack in the cloud is more as the cloud environment utilizes shared infrastructure pool powered by virtualization as its backbone. Example: DDoS is a common attack in all the networking environments. But the DDoS attack can cause more damage to cloud environment than any other environment, because tracing the source of the attack is difficult as the attacker itself can be a virtualized machine which can appear and disappear easily, quickly in the attack environment through automated scripts. Thus it will lead to ultimate difficulty in tracing the source of attack. The recent DDoS attack on a web hosting Service Company “Bit Bucket” running on Amazon Cloud let the service down for more than 19 hours [11]. As any other environment cloud is also prone to security problems like Data Loss, Phishing, Password Cracking, Downtimes, Botnets and other malwares. The Data Loss not a just a common problem it is more specific to cloud, hence it is dealt separately in Data Security section.

Level	User	Requirements	Threats
Application Level (SaaS)	Person or Organization subscribed to a cloud Provider	Service Availability Communication Protection Access Control Privacy Data protection	Privacy Breach Traffic Flow Analysis Session Hijacking Exposure in network Impersonation
Virtual Level (PaaS, IaaS)	Person or Organization that deploy software on the cloud infrastructure	Virtual Cloud protection Secure images Cloud Management Security Application Security	Connection Flooding Programming flaws Software modification Software Modification Software interruption DDoS
Physical Level	Person or Organization owns the infrastructure	Hardware Security Hardware reliability Network protection Network resource Protection	Network Attacks Hardware theft Natural Disasters Hardware Modification DDoS

4. CLOUD SPECIFIC SECURITY ISSUES

A Vulnerability can be qualified as Cloud Specific [3] if it is intrinsic to or prevalent in a core cloud computing technology has its roots in one of NIST’s essential cloud characteristics is caused when cloud innovations make tried-and tested security controls difficult or impossible to implement or is prevalent in established state-of-the-art cloud offerings.

Some of the security issues which can be deemed as Cloud Specific are explained here.

A. Accountability

The Accountability is an issue of holding the right person responsible for the security lapse, if any happened. Since the cloud service activities are not transparent among the stake holders, no one can be held responsible for certain security lapses or there will be confusion in holding some body accountable. Hence a clear demarcation is necessary between the stake holder’s role in the cloud services as well as the transparency. Establishing the demarcation and transparency is difficult as the cloud services spans across layers. The cloud services architecture is not standardized, every new application is unique on its own, and hence

delegating accountability rests with the Cloud Service Provider. Even if the accountability is delegated by service provider, establishing accountability under the incidence of security lapse rest with all the stake holders who shares the accountability. For example if a web hosting service company hosts its service on cloud based infrastructure services company like Amazon, if a client of web hosting service company loses data, who will be held accountable, Infrastructure Provider or Web Provider. It is difficult to establish accountability. Hence there is an utmost necessity of solving accountability problems in cloud.

B. No Security Perimeter

The Cloud is mostly powered by virtualized shared pool of resources. The computing requirement of the cloud consumers is mostly limited to Virtual Machines VMs. When the client access the cloud, the client will be provided a new VM instance which will be mostly dedicated to that client, also the new VMs will be provisioned on dynamic scaling. But the Cloud provides little control over physical or network location of VM instances. Thus providing perimeter level security like IDS, Firewall etc... is difficult for hosted services on cloud Infrastructure. Hence network access must be controlled on a host by host basis.

C. Larger Attack Surface

The cloud provides a larger attack surface when compared to the traditional network environment. Since the VMs launched in a cloud environment have little control over control over physical or network location, the placement of VMs is not limited by boundary. Hence an infected VM can be placed anywhere in the cloud, which leads to limitless attack surface. The Botnets spread will lead to severe DDoS attack in Cloud due to its larger attack surface.

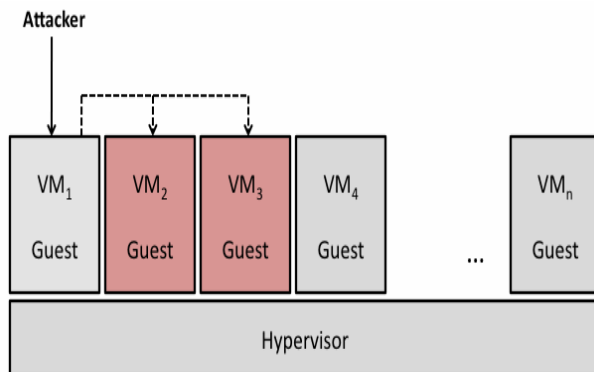


Fig. 5. Infected VMs spread

D. NEW SIDE CHANNELS

The Side Channels are inevitable part of the system, whose unintentional leakage can be exploited by the attacker to

extract some private information like cryptographic keys. In the cloud, utilization of shared physical hardware by the VMs, introduces new type of side channels like CPU data cache, CPU instruction cache, CPU Branch Prediction etc... The VMs who share the hardware can exploit these side channels to steal the private information of neighbor VM hosted on the same physical machine. For example the CPU data cache is shared among the entire guest VMs running on the same physical machine. If any attacker VM is located on the same machine, it can access the data cache trace to predict the cryptographic keys used by victim VMs.

E. LACK OF AUDITABILITY

Most of the companies will have audit requirements in connection with their nature of business and client base. In general, any organization will have definite audit process to meet their quality requirements. There is no exception to business on cloud too; they also need to satisfy their audit process. But the problem with the cloud is “only cloud provider has access to full network traffic, hypervisor logs, and physical machine data.” To provide complete auditability features, the cloud needs mutual auditability features. It should enable the cloud provider the ability to audit potentially malicious or infected client VMs. Also, it should enable the cloud consume the ability to audit cloud provider environment.

F. EDoS

EDoS – Economic Denial of Service, the cloud is more vulnerable to traditional DDoS which can be transformed into EDoS, targeting the dynamic scaling feature of Cloud. The Attack can be initiated by malicious insider as well as rival outsider. In general, the DDoS will target the performance of the victim service by overwhelming the service access by excessive illegitimate requests that the server can handle. Thereby renders the legitimate users devoid of service. But the service hosted on the cloud can compensate the attack requests by dynamically scaling excessive resources, which will avoid performance degradation, but will target the economic viability of the service as the attack scales up the resource consumption there by service provider has to pay more for the host cloud. Thus in cloud environment the DDoS attack will be transformed into EDoS attack.

G. REGULATORY COMPLIANCE

As the cloud infrastructure transcends the geographical boundaries, the hosted services have to comply with the regulatory compliance of the region under which the service is running. This is a common issue for any hosted services, but with respect to cloud hosted services, it introduces new problem. A service hosted in the one region may have their VM instance running in other region, because the VM will be launched by the cloud provider, whereas the service

provider will have little or no control over the placement of instances, which may lead to regulatory compliance issue, if the service breaches the regulatory compliance of the region where the VM is alive. It is difficult to meet regulatory compliances for the critical services like military operations of a country. The cloud must ensure regulatory compliance to host critical services.

H. DATA SECURITY

The data security is the most important security to be addressed in cloud. The cloud adoption apprehensions are mostly because of the security lapse witnessed in data security. In the cloud, the data will be in various states. The data may be in transit, it may be at rest or at processing. In

cloud, ensuring data security means, provision of security to data in all the states. The data security should ensure confidentiality, availability and integrity of the data in all the states. The availability of the data can be provided by redundancy in all the states. The integrity and confidentiality can be addressed by SSL, Homographic encryption, Symmetric encryption etc... at various states of data.

V. COUNTER MEASURES

The analysis of various counter measures proposed, practiced and experimented by various researchers and practitioners for different types of security vulnerabilities and attacks is detailed in the Table II. The table gives the summary of various approaches to secure cloud.

Table II. Summary of Countermeasures

Approaches	Focus	Methodology	Distributed approach	Learning ability	Balances the workload	Tolerance to failure	Time response	Scalability
Distributed Cloud Intrusion Detection Model [12]	DDoS attack and cross site scripting	Intrusion Detection system	Yes	Yes	Yes	No	Real Time	Yes
A New Trusted and Collaborative Agent Based Approach [13]	SQL injection attack, Cross site scripting, DDoS.	Trust and Authentication	Yes	No	No	No	Real Time	No
Implementing Trust in Cloud Infrastructures [14]	DDoS attack	Trust based	No	No	No	Yes	Real Time	Yes
CBF: A Packet Filtering Method for DDoS Attack Defense [15]	Distributed Denial-of-Service attack	Packet Filtering	Yes	Yes	Yes	No	Real Time	Yes
Defend Against DDoS Attack with VMM [16]	DDoS attacks	Traffic monitoring	No	No	Yes	No	-	Yes
EDoS-Shield [17]	EDoS attack	Virtual firewall and authentication	No	Yes	No	No	Real Time	Yes
Cloud Traceback [18]	HTTP and XML based DoS Attack	Packet marking and Traceback	Yes	Yes	Yes	Yes	Real Time	Yes
sPoW: On-Demand Cloud-based eDDoS Mitigation [19]	EDoS attack	Packet Filtering	Yes	Yes	Yes	No	Real Time	Yes
A Layered Security Approach for Cloud Computing [20]	DDoS attack	Security architecture	No	No	Yes	No	-	No
Addressing cloud computing security issues [8]	Common	Trust, cryptography and certificate.	Yes	No	No	No	-	No

5. CONCLUSION

Cloud Computing provides a wide range of services across various levels viz... Application, Platform and Infrastructure, since the range of services spans different levels; it introduces complexity in security requirements. Also the cloud's innovative services coupled with shared infrastructure introduce new range of security threats, vulnerabilities as well as security requirements to counter the attacks. The paper explored the security issues which are cloud specific and analyzed the counter measures available. For wide adoption of cloud computing, the security apprehensions need to be weeded away from the minds of cloud service consumers.

REFERENCES

- [1] Alexander Lenk, Markus Klems, Jens Nimis, Stefan Tai, Thomas Sandholm, "What's Inside the Cloud? An Architectural Map of the Cloud" In: International Conference on Software Engineering, Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, 2009, p.23-31, ISBN:978-1-4244-3713-9.
- [2] A Taxonomy and Survey of Cloud Computing System", Fifth International Joint Conference on INC, IMS and IDC.
- [3] Bernd Grobauer, Tobias Walloschek, and Elmar Stöcker "Understanding Cloud Computing Vulnerabilities" Cloud Computing, Copublished by The IEEE Computer And Reliability Societies.
- [4] Krešimir Popović, Željko Hocenski "Cloud computing security issues and challenges", MIPRO 2010, May 24-28, 2010, Opatija, Croatia.
- [5] Paul Wooley, Tyco Electronics, "Identifying Cloud Computing Security Risks".
- [6] V Venkateswara Rao, G. Suresh Kumar, Azam Khan, S Santhi Priya, "Threats and Remedies in Cloud.
- [7] A Survey on Cloud Computing Security, Challenges and Threats", Journal of Current Computer Science and Technology Vol. 1 Issue 4[2011]101-106.
- [8] Dimitrios Zissis, Dimitrios Lekkas, "Addressing cloud computing security issues", Future Generation Computer Systems.
- [9] "Critical Areas of Focus in Cloud Computing", Cloud Security Alliance, December 2009.
- [10] Ketki Arora, Krishan Kumar, And Monika Sachdeva, "Impact analysis of DDoS Attack", International Journal on Computer Science and Engineering (IJCSSE)- Vol. 3 No. 2 Feb 2011.
- [11] Metz C, "DDoS attack rains down on Amazon Cloud", http://www.theregister.co.uk/2009/10/05/amazon_bitbucket_outage
- [12] Irfan Gul, M. Hussain "Distributed Cloud Intrusion Detection Model" International Journal of Advanced Science and Technology Vol. 34, September, 2011.
- [13] Shantanu Pal, Sunirmal Khatua, Nabendu Chaki, Sugata Sanyal, "A New Trusted and Collaborative Agent Based Approach for Ensuring Cloud Security Implementing Trust in Cloud Infrastructures".
- [14] Qi Chen, Wenmin Lin, Wanchun Dou, Shui Yu "CBF A Packet Filtering Method for DDoS Attack Defense in Cloud Environment"
- [15] Siqin Zhao, Kang Chen, Weimin Zheng, "Defend Against Denial of Service Attack with VMM", Eighth International Conference on Grid and Cooperative Computing.
- [16] Mohammed H. Sqalli Fahd Al-Haidari Khaled Salah, "EDoS-Shield - A Two-Steps Mitigation Technique against EDoS Attacks in Cloud Computing "Fourth IEEE International Conference on Utility and Cloud Computing.
- [17] Ashley Chonka, Yang Xiang n, Wanlei Zhou, Alessio Bonti, "Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks", Journal of Network and Computer Applications 34 (2011) 1097-1107.
- [18] Soon Hin Khor Akihiro Nakao, "sPow On-Demand Cloud-based eDDoS Mitigation Mechanism".
- [19] Mehmet Yildiz, Jemal Abawajy, Tuncay Ercan and Andrew Bernoth, "A Layered Security Approach for Cloud Computing Infrastructure", 10th International Symposium on Pervasive Systems, Algorithms, and Networks.
- [20] Thomas Ristenpart, Eran Tromer, Hovav Shacham, Stefan Savage, "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds".
- [21] Kazi Zunnurhain, Susan V. Vrbsky, "Security in cloud computing",
- [22] Meiko Jensen, Jörg Schwenk, Nils Gruschka, Luigi Lo Iacono, "On Technical Security Issues in Cloud Computing", IEEE International Conference on Cloud Computing.
- [23] Nils Gruschka and Luigi Lo Iacono, "Vulnerable Cloud: SOAP Message Security Validation Revisited", IEEE International Conference on Web Services.
- [24] <http://www.bankinfosecurity.com/p-survey-cloud-security-2012>
- [25] "Overcoming the Apprehension of Cloud Computing: Results from the 2012 Cloud Computing Security Survey ", Information Security Media Group, 2012.

Single Document Summarization Using TF/IDF Technique

Nandini Anand¹, Mayank Baheti², Prakhar Rastogi³, Atul Kumar Srivastava⁴

^{1,2,3}Student, B.tech (E & T,) ⁴Asst. Professor, AITTM

^{1,2,3}AITTM, Amity University ⁴Amity University

¹anandnandini24@gmail.com, ²mayankbaheti25@gmail.com, ³prakhar.supermintexports@gmail.com

Abstract: Summarization is the art of abstracting key content from one or more information sources that has become an integral part of everyday life. People keep abreast of world affairs by listening to news bites. They base investment decisions on stock market updates. They even go to movies largely on the basis of reviews they've seen. With summaries, they can make effective decisions in less time.

With a large volume of text documents, presenting the user with a summary of each document greatly facilitates the task of finding the desired documents. Text search and summarization are the two essential technologies that complement each other. Text search engines return a set of documents that seem to be relevant to the user's query, and text enable quick examinations through the returned documents.

In general, automatic document summarization takes a source document (or source documents) as input, extracts the essence of the source, and presents a well-formed summary to the user.

Keywords: Automatic Summarization, TF/IDF, Visual Basics

1. INTRODUCTION

Mani and Maybury defined automatic document summarization as the process of distilling the most important information from a source to produce an abridged version for a particular user (or users) and task (or tasks). The process can be decomposed into three phases: analysis, transformation and synthesis. The analysis phase analyzes the input document and selects a few salient features. The transformation phase transforms the results of analysis into a summary corresponding to users' needs. In the overall process, compression rate, which is defined as the ratio between the length of the summary and that of the original, is an important factor that influences the quality of the summary. While the compression rate increases, the summary will be more copious, relatively, more insignificant information is contained. [3]

The goal of *automatic summarization* is to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or application's needs.

There are different types of summaries depending what the summarization program focuses on to make the summary of the text, for example *generic summaries* or *query relevant summaries* (sometimes called *query-based summaries*).

A summary can also be indicative, informative, or critical:

- *Indicative* summaries follow the classical information retrieval approach, i.e. they provide enough content to alert users to relevant sources, which users can then read in more depth.
- *Informative* summaries act as substitutes for the source, mainly by assembling relevant or novel factual information in a concise structure.
- *Critical* summaries (or reviews), besides containing an informative gist, incorporate opinion statements on content. They add value by bringing expertise to bear that is not available from the source alone. [4]

A summary can also be generic or user-focused:

- *Generic* summaries address a broad community; there is no focus on special needs because the summarizer is not targeting any particular group.
- *User-focused* summaries, in contrast, are tailored to the specific needs of an individual or a particular group (children, for example).

Until recently, generic summaries were more popular, but with the prevalence of full-text searching and personalized information filtering, user-focused summaries are gaining importance. Many tools support both user-focused and generic summarization.

Summarization systems are able to create both query relevant text summaries and generic machine-generated summaries depending on what the user needs. Summarization of multimedia documents, e.g. pictures, videos or movies, is also possible.

Some systems will generate a summary based on a single source document, while others can use multiple source

documents (for example, a cluster of news stories on the same topic). These systems are known as *multi-document summarization* systems. [5]

2. METHOD OF ARCHITECTURE

Automatic summarization involves reducing a text document or a larger corpus of multiple documents into a short set of words or paragraph that conveys the main meaning of the text. This can be done by two methods: extraction or abstraction.

Extractive methods work by selecting a subset of existing words, phrases, or sentences in the original text to form the summary. In contrast, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a human might generate. Such a summary might contain words not explicitly present in the original. The state-of-the-art abstractive methods are still quite weak, so most research has focused on extractive methods, and this is what we will cover. Two particular types of summarization often addressed in the literature are key phrase extraction, where the goal is to select individual words or phrases to "tag" a document, and document summarization, where the goal is to select whole sentences to create a short paragraph summary. [5]

Extraction approaches are easy to adapt to larger sources. Because they are limited to the extraction of passages, sentences, or phrases, however, the resulting summaries may be incoherent. That's why, most of the summarization work done till date is based on extraction of sentences from the original document. The sentence extraction techniques compute score for each sentence based on features such as position of sentence in the document [Baxendale 1958; Edmundson 1969], word or phrase frequency [Luhn 1958], key phrases (terms which indicate the importance of the sentence towards summary. The architecture of extraction technique is shown in figure1. [4]

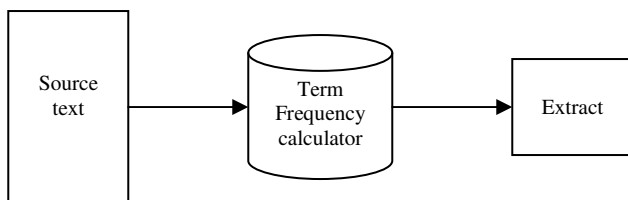


Fig. 1. Architecture of Extraction

Sentence extraction is a technique used for automatic summarization of a text. In this shallow approach, statistical heuristics are used to identify the most salient sentences of a text. In short, "sentence extraction" works as a filter which allows only important sentences to pass.

Usually, a combination of heuristics is used to determine the most important sentences within the document. Each heuristic assigns a score to the sentence. After all heuristics have been applied, the highest-scoring sentences are included in the summary. The individual heuristics are weighted according to their importance.

Most of the summarization work done till date is based on extraction of sentences from the original document. Many basic methods are there like based on font size, style or by paragraph selection, different size of headings and sub-headings, etc but these has some limitations. We follow a better approach, by using tf/idf technique.

3. PROPOSED APPROACH

In this paper we present an automatic summarization system, which generates a summary for a given input document. Our system is based on identification and extraction of important sentences in the input document. The architecture of our summarization system has both text analysis component and summary generation component.

The text analysis component is based on syntactic analysis, followed by a component which identifies the rank for each sentence. As a part of summarization, we try to identify the important sentences which represent the document. This involves considerable amount of text analysis. We assume that the input document can only be of text document format (i.e. notepad) hence first we manually convert any document (e.g. PDF, MS Word, post-script, HTML...) into text. [2]

1. Text Normalization

The text normalization is a rule based component which removes the unimportant objects like figures, tables, handling of non-standard words like web URL's and emails but this is done manually in our system.

2. Sentence Marker

This module divides the document into sentences. It considers dot (.) as the end-of-sentence and marks the boundary of sentences. With an exception that we are considering there would be no abbreviations and short forms used in the text.

3. Feature Extraction

- Firstly, all stop words are removed from the text.

A *stop word* is a term you specify to exclude from your database because it occurs too frequently or because of its unimportance to the document content. Typical stop words include 'a', 'an', 'and', 'the', 'but', and so on.

- Then term frequency (tf) of each word in the document is calculated.

Term frequency is the raw frequency of a term in a document, i.e. the number of times that term occurs in the document. [2]

Summary generation component include tasks such as calculating the score for each sentence, selecting the sentences with high score, and generating a selected pool of sentences.

1. Sentence Ranking

Once the term frequency for each word is extracted, the score for each sentence is generated i.e. the sum of term frequency of individual words that occur in that sentence.

$$\text{Weight}(w) = \sum \text{Weight}(w, l)$$

Where w denotes the sentence number and l denotes the word that occurred in the sentence.

2. Sentence Selection

After the sentences are scored, we need to select the sentences that make good summary. Our strategy is to pick the top N sentences to generate the summary. The selection of sentence is dependent upon the type of the summary requested.

Sentence selection module will give a set of sentences which satisfies the user criteria, i.e. length of the summary 10%, 20%, 30%, and so on of the original document. [2]

4. ALGORITHM

The algorithm our proposed model is as follows:

- Load the document into memory.
- Load stop words list into the memory
- Split the text into sentences and create sentence list.
- Remove the stop words from the sentences in sentence list.
- Calculate the **term frequency** (the raw frequency of a term in a document, i.e. the number of times that term occurs in the document) of each word in the document
- Calculate the **weight** of each sentence, which is the sum of the term frequency of all words that occur in that sentence.
- Then sort these sentences according to their weight, keeping the highest at the top.
- Now pick the top N sentences, as per the user's requirement.

5. RESULT

Algorithm is explained by the help of an example as follows:

1. Take a sample document

Adolescence is a transitional stage of physical and psychological human development generally occurring between puberty and legal adulthood. The period of adolescence is most closely associated with the teenage years, although its physical, psychological and cultural expressions can begin earlier and end later. Adolescence is viewed as a transitional period between childhood and adulthood whose cultural purpose is the preparation of children for adult roles. Such milestones include driving a vehicle, serving in the armed forces or on a jury, voting, entering into contracts, completing certain levels of education, and marriage. Adolescence is usually accompanied by an increased independence allowed by the parents or legal guardians and less supervision as compared to pre-adolescence. In popular culture, many adolescent characteristics are attributed to physical changes and raging hormones. In studying adolescent development, adolescence can be defined biologically, as the physical transition marked by the onset of puberty and the termination of physical growth; cognitively, as changes in the ability to think abstractly and multi-dimensionally; or socially, as a period of preparation for adult roles.

2. Break the text into sentences

1.	Adolescence is a transitional stage of physical and psychological human development generally occurring between puberty and legal adulthood.
2.	The period of adolescence is most closely associated with the teenage years, although its physical, psychological and cultural expressions can begin earlier and end later.
3.	Adolescence is viewed as a transitional period between childhood and adulthood whose cultural purpose is the preparation of children for adult roles.
4.	Such milestones include driving a vehicle, serving in the armed forces or on a jury, voting, entering into contracts, completing certain levels of education, and marriage.
5.	Adolescence is usually accompanied by an increased independence allowed by the parents or legal guardians and less supervision as compared to pre-adolescence.
6.	In popular culture, many adolescent characteristics are attributed to physical changes and raging hormones.
7.	In studying adolescent development, adolescence can be defined biologically, as the physical transition marked by the onset of puberty and the termination of physical growth; cognitively, as changes in the ability to think abstractly and multi-dimensionally; or socially, as a period of preparation for adult roles.

3. Remove the stop words from the sentences

1.	Adolescence transitional stage physical psychological human development generally occurring puberty legal adulthood
2.	Period adolescence closely associated teenage years physical psychological cultural expressions earlier
3.	Adolescence viewed transitional period childhood adulthood cultural purpose preparation children adult roles
4.	Milestones include driving vehicle serving armed forces jury voting entering contracts completing levels education marriage
5.	Adolescence accompanied increased independence allowed parents legal guardians less supervision compared pre-adolescence
6.	Popular culture adolescent characteristics attributed physical changes raging hormones
7.	Studying adolescent development adolescence defined biologically physical transition marked onset puberty termination physical growth cognitively changes ability think abstractly multi-dimensionally socially period preparation adult roles

4. Now calculate the term frequency of each word in the text

5. Calculate the weight of each sentence

1.	Adolescence is a transitional stage of physical and psychological human development generally occurring between puberty and legal adulthood. (25)
2.	The period of adolescence is most closely associated with the teenage years, although its physical, psychological and cultural expressions can begin earlier and end later. (22)
3.	Adolescence is viewed as a transitional period between childhood and adulthood whose cultural purpose is the preparation of children for adult roles. (24)
4.	Such milestones include driving a vehicle, serving in the armed forces or on a jury, voting, entering into contracts, completing certain levels of education, and marriage. (15)
5.	Adolescence is usually accompanied by an increased independence allowed by the parents or legal guardians and less supervision as compared to pre-adolescence. (16)
6.	In popular culture, many adolescent characteristics are attributed to physical changes and raging hormones. (15)
7.	In studying adolescent development, adolescence can be defined biologically, as the physical transition marked by the onset of puberty and the termination of physical growth; cognitively, as changes in the ability to think abstractly and multi-dimensionally; or socially, as a period of preparation for adult roles. (45)

6. Sort these sentences according to their weight, in descending order

1.	In studying adolescent development, adolescence can be defined biologically, as the physical transition marked by the onset of puberty and the termination of physical growth; cognitively, as changes in the ability to think abstractly and multi-dimensionally; or socially, as a period of preparation for adult roles. (45)
2.	Adolescence is a transitional stage of physical and psychological human development generally occurring between puberty and legal adulthood. (25)
3.	Adolescence is viewed as a transitional period between childhood and adulthood whose cultural purpose is the preparation of children for adult roles. (24)
4.	The period of adolescence is most closely associated with the teenage years, although its physical, psychological and cultural expressions can begin earlier and end later. (22)
5.	Adolescence is usually accompanied by an increased independence allowed by the parents or legal guardians and less supervision as compared to pre-adolescence. (16)
6.	Such milestones include driving a vehicle, serving in the armed forces or on a jury, voting, entering into contracts, completing certain levels of education, and marriage. (15)
7.	In popular culture, many adolescent characteristics are attributed to physical changes and raging hormones. (15)

7. Now pick top N sentences as per the user's requirement

In studying adolescent development, adolescence can be defined biologically, as the physical transition marked by the onset of puberty and the termination of physical growth; cognitively, as changes in the ability to think abstractly and multi-dimensionally; or socially, as a period of preparation for adult roles. Adolescence is a transitional stage of physical and psychological human development generally occurring between puberty and legal adulthood. Adolescence is viewed as a transitional period between childhood and adulthood whose cultural purpose is the preparation of children for adult roles.
--

6. WORKING OF AUTOMATIC SUMMARIZER

Coding of the system is done in *visual basics* language, and the system is named as Automatic Summarizer. After installing the set-up successfully, a start-up window appears on the screen, as shown in figure2. The main window displays different icons like summarization, advance, help and window. First, click on load file and load the text

document, as shown in figure4. Then click on do summary, to choose from the given options, as shown in figure5. Summary is generated as per the user's requirement, as shown in figure6. Then click on sentence ranking to see the sentence list along with their weights, as shown in figure7. Figure8 shows the cascade view of summary generation having the original text, sentence list and summary.



Fig. 2. Start-up window of the software

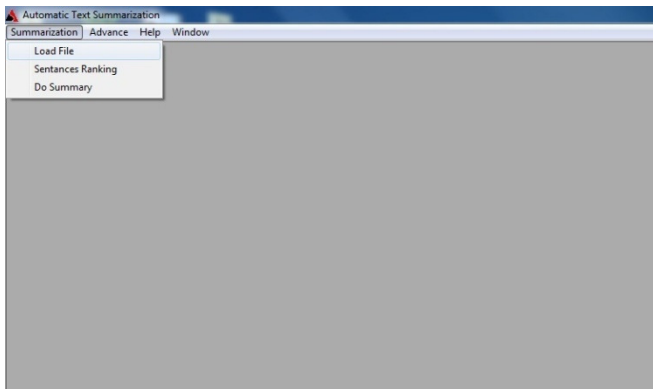


Fig. 3. Summarization Menu

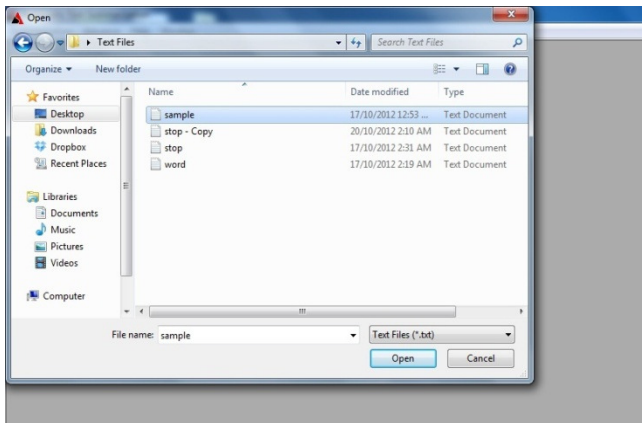


Fig. 4. Load file

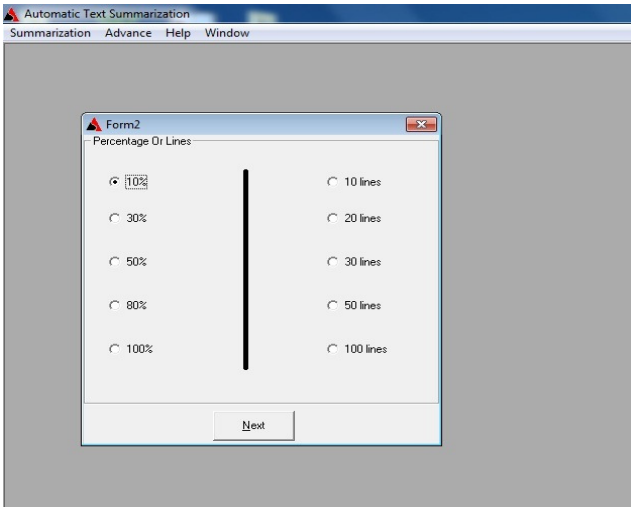


Fig. 5. Summary Selection

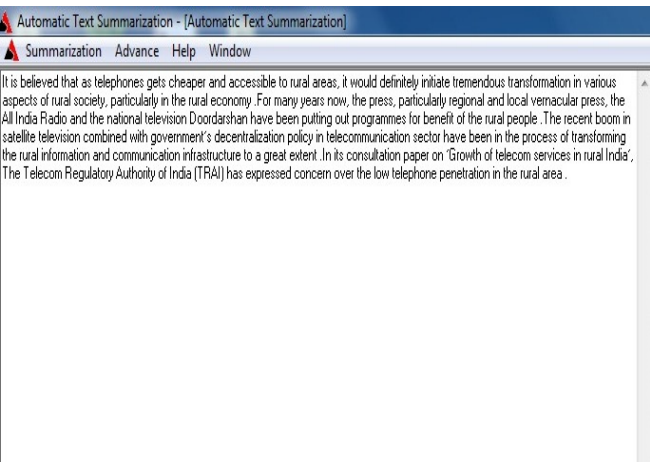


Fig. 6. Generated summary

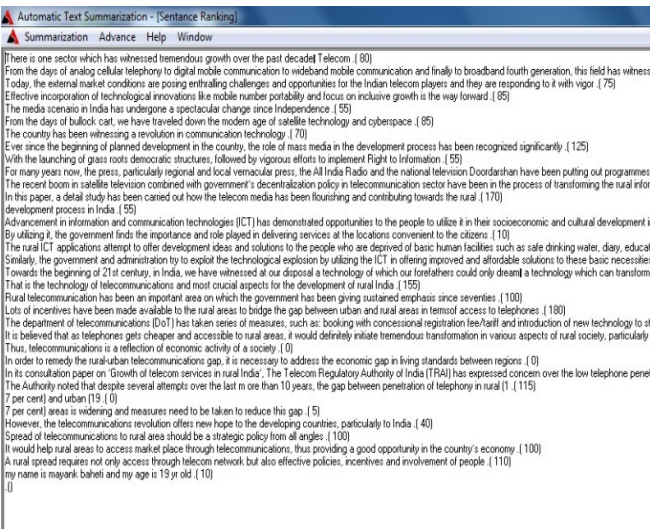


Fig 7. Sentence List with their weight

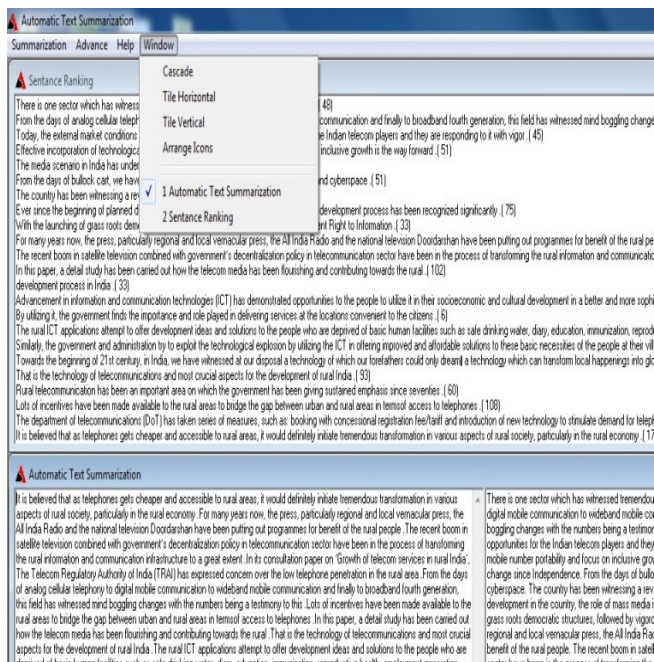


Fig. 8. Cascade View

7. CONCLUSION

The process of summarizing documents is becoming increasingly important in the light of recent advances in document creation/distribution technology, and the resulting influx of large numbers of documents in every day life.

In this paper we presented a sentence extraction based single document summarization system that extracts the most relevant sentences from the source document to form a generic summary. Our proposed method shows comparable performance to other basic methods.

We used shallow text processing approaches as opposed to semantic approaches related to natural language processing. We presented a detailed architecture and internal working of our system. We defined a ranking function which ranks each sentence as a linear combination of the sentence features. The experiments showed that the summary generated is coherent and the selected features are really helpful in extracting the important information in the document.

The system can be used in many applications, as follows:

- By using this system people can save a lot of time, as instead of reading the whole document they will just read the summary to get the gist of the whole document.
- It can be used to produce summary reports for text contents by processing documents, web pages, hyperlinks, e-mail messages and files.
- It can provide summary of PDF files, MS Word, postscript, HTML document, or any other document.
- By integrating with Internet Explorer it can also generate summaries of arbitrary texts, including web pages.
- By doing some addition to the program we can also get target specifications in terms of number of sentences, some percentages of size and number of words, etc. of the document.
- This also has scope in web-mining and in question-answering on web-search.

REFERENCES

- [1] Mani. 2001. Automatic Summarization. John Benjamins Pub. Co.
- [2] J Jagadeesh, P Pingali, V Varma - International Institute of, 2005 - web2py.iit.ac.in
- [3] RM Aliguliyev - Вычислительные технологии, 2007 - ict.nsc.ru
- [4] U Hahn, I Mani - Computer, 2000 - icccexplore.ieee.org
- [5] A Nenkova, R Passonneau - Proceedings of HLT-NAACL, 2004 - acl.ldc.upenn.edu
- [6] Hirao, Y. Sasaki, H. Isozaki, and E. Maeda. 2002. Ntt's text summarization system for DUC-2002. In DUC 2002: Workshop on Text Summarization, July 11-12, 2002, Philadelphia, PA, USA.
- [7] http://www.worldscientific.com/Y_Matsuo, M Ishizuka - International Journal on Artificial Intelligence ..., 2004
- [8] E Hovy, CY Lin - Proceedings of a workshop on held at Baltimore..., 1998 - dl.acm.org
- [9] E Hovy, D Marcu - The Oxford Handbook of computational ..., 2005 - coli.uni-saarland.de
- [10] J Goldstein, V Mittal, J Carbonell... - ... Automatic summarization- ..., 2000 - dl.acm.org

Language Morphology and Search Engine Performance

Nargis Parveen¹, Mohd Athar²

^{1,2}Research Scholar, Shri Venkateshwara University, Gajraula, India

¹nargis.parveen@gmail.com, ²m.atharjmi@gmail.com

Abstract: The morphological, structural and grammar related issues of languages are generally ignored by web searchers during their query formulation and searching. In fact, these factors can be very important in improving the performance of search engines. In this paper, we will show an effort to highlight three factors. Our results show that the performance of the search engines is affected by these factors. The query term ambiguity may sometimes drastically reduce the relevancy of a search engine. Hence it is to be dealt with properly through automated algorithm for disambiguation. Overall, the search engines can be made more users friendly and productive by appropriately handling these issues.

Index Terms: Web Ambiguity, search engine, morphology.

1. INTRODUCTION

The term 'morphology' refers to the study of the internal structure of words, and of the systematic form-meaning correspondences between words. Morphology is the study of the structure of words. The structure of words can also be studied to show how the meaning of a given morpheme, or its relation to the rest of the word, varies from one complex word to another. Consider how sun works in the following words: sunbeam, sunburn, sundial, sunflower, sunglasses, sunlight, sunrise, and sun-spot (scientific sense), and sun-spot (tourist sense), and suntan. Inflection does not really yield "new" words, but alters the form of existing ones for specific reasons of grammar. Derivation, on the other hand, does lead to the creation of new words.

Morphology is the field of linguistic which studies word structure and formation [12]. It is composed of inflectional morphology and derivational morphology [13, 14]. Inflection is defined as the use of morphological methods to form inflectional word forms from a lexeme. Inflectional word forms indicate grammatical relations between words. Derivational morphology is concerned with the derivation of new words from other words using derivational affixes. Compounding is another method to form new words. A compound word (or a compound) is defined as a word formed from two or more words written together. The component words are themselves independent words (free morphemes).

A morpheme is a smallest unit of a language which has a meaning. Morphemes are classified into free morphemes and bound morphemes [14, 15]. Free morphemes appear as independent words. e.g. In English, {red}, {house} and {when} are free morphemes. Bound morphemes do not constitute independent words, but are attached to other morphemes or words. Bound morphemes are also called affixes. Morphological structure of English language has a great impact on the performance of the search engines. In this study we have focused on three factors of language morphology that can change or modify a web query i.e. query with root word, query with different synonym and query with various senses.

Methods of Evaluation of Search Engines

There are following methods which are used for the evaluation of search engine:

Precision (P): is the fraction of retrieval documents that are relevant. A high precision means that everything returned was a relevant result, but one might not have found all the relevant items (which would imply low recall). There are variations in the ways of the precision is calculated. TREC almost always uses binary relevance judgments—"either a document is relevant to a query or it is not" [16]. Chu & Rosenthal [17] used a three-level relevance score (relevant, somewhat relevant, and irrelevant) while Gordon and Pathak [18] used a four-level relevance judgment (highly relevant, somewhat relevant, somewhat irrelevant, and highly irrelevant).

RECALL (R): It is the fraction of relevant documents that are retrieved. A high recall means we haven't missed anything but we may have a lot of useless results to sift through (which would imply low precision). But Recall is a difficult measure to calculate because it requires the knowledge of the total number of relevant items in the collection. Chu & Rosenthal's Web search engine study omitted recall as an evaluation measure because they consider it "impossible to assume how many relevant items are there for a particular query in the huge and ever changing Web systems" [17]. Based on the documents retrieved by a

search engine (relevant, non relevant), Table 1 below shows the method of computations of precision and recall.

TABLE 1: Precision and Recall Computation Table

	Relevant	Non-relevant
Retrieved	True positives (tp) - Correct result	False positives (fp)- Unexpected result
Not retrieved	False negatives (fn) - Missing result	True negatives (tn) - Correct absence of

The precision and recall can be calculated by the formula shown below: Precision = $tp / (tp+fp)$
Recall = $tp / (tp+fn)$

Where tp is retrieved relevant result, And fp is retrieved non relevant result, And fn is missing result (i.e. relevant but not retrieved)

Mean Average Precision (Map):

Most standard among the TREC community is Mean Average Precision (MAP), which provides a single-figure measure of quality across recall levels. Among evaluation measures, MAP has been shown to have especially good discrimination and stability. For a single information need, Average Precision is the average of the precision value obtained for the set of top k documents existing after each relevant document is retrieved, and this value is then averaged over information needs.

MAP = Average Precision/ No. of queries

When a relevant document is not retrieved at all, the precision value in the above equation is taken to be 0. Why these methods are used?

These methods are used because the users always want see some documents, and can be assumed to have a certain tolerance for seeing some false positives providing that they get some useful information. The measure of precision and recall concentrate the evaluation on the return of true positive, asking what percentage of the relevant documents have been found and how many false positive have also been returned.

Evaluation Methodology

The U.S. National Institute of Standards and Technology (NIST) have run a large IR test based evaluation series since 1992. Within this framework, there have been many tracks over a range of different test collections, but the best known test collections are the ones used for the TREC Ad Hoc track during the first eight TREC evaluations between 1992 and 1999. TRECs 6 through 8 provide 150 information needs

over about 528,000 newswire and Foreign Broadcast Information Service articles. In this work, we have framed the queries based on the TREC pattern and also from the web search engine's log. So our set of test queries used for the evaluation of search engines in this study have a good mix of standard TREC queries and actual user queries from the search engine's log.

Human Relevance Judgments:

It is one of the important issues in performance evaluation of search engines is that whenever human relevance judgment is used, there is a variation in who makes the judgments. TREC leaves relevance judgments to experts or to a panel of experts (Voorchees & Harman, 2001) [16]. However some other researchers (e.g. Chu and Rosenthal, 1996) used human relevance judgment made by researchers themselves. Gordon and Pathak [18] emphasized that relevance judgments can only be made by individual with the original information need. In this study, the human relevance judgments have been done using a mix of the approaches followed by Voorchees et.al (2001) and Chu et.al. (1996).

Precision:

There are variations in the ways how precision is calculated. In this study, the precision is calculated on the binary relevance judgment approach followed by TREC -"either a document is relevant to a query or it is not" [16].

Recall:

Chu & Rosenthal's [17] Web search engine study omitted recall as an evaluation measure because they consider it "impossible to assume how many relevant items there are for a particular query in the huge and ever changing Web systems". In this study too we have omitted the recall as an evaluation measure for the similar reasons.

The computation of precision has been done as follows: Suppose an IR system returns 8 relevant documents and 10 non-relevant documents. There are a total of 20 relevant documents in the collection.

$$\begin{aligned}
 tp \text{ (true positive)} &= 8 \\
 fp \text{ (false positive)} &= 10 \\
 fn \text{ (false negative)} &= 20-8=12 \\
 \text{Precision} &= tp / (tp+fp) = 8 / (8+10) \\
 &= 8/18 = 0.44
 \end{aligned}$$

Average Precision = sum of all precision/ No. of queries
Mean Average Precision = av. precision/ No. of queries

Factors Affecting Performance of Search Engines

The information retrieval on the web in any language faces numerous challenges. Besides all the technical factors the

grammatical and morphological structure of the language is one of the critical factors that can affect the performance of the information retrieval system on the web.

Root Word of The Keywords:

In English prefixes and suffixes (collectively called affixes) are normally used (e.g. s, es, dis, ness, ing etc.) with morpheme (root word) and new words are constructed. These new words are called morphological variants of the stem.

For ex.: increase + ing = increasing, or dis + able = disable. Or happy + ness = happiness.

While searching on the web the query terms given by the users may not be in root form. As there is no restriction/help about how to choose or select the query term, same query may be formed with different morphological variations of its terms. This may lead to variation of results and the relevancy of results by search engines. To analyze this, we took a real time test of Google search engine using a set of 20 web queries (as per the discussion in the previous section). These queries are listed in table 2, and to properly analyze the result each query has been written twice - with root words and without root words.

TABLE 2: Test Query Set For Root Word Analysis

Query with root word	Query without root word
Civil Service exam	Civil service examination
Mercury level in bird	Mercury levels in birds
water waste in India	water wastage in India
Fund and grants institution	Funding and grants institution
beds sharing with children	beds sharing with children's
mercury levels is increase	mercury levels is increasing
The temperature is decrease	The temperature is decreasing
Native language of India	Native languages of India
merit of democracy	merits of democracy
Use of computer	Uses of computer
demerit of democracy	demerits of democracy
advantage of mobile phones	advantages of mobile phones
disadvantage of mobile phones	disadvantages of mobile phones
Imagine power	Imagination power
power of battery	power of batteries
liberty of information act forms	liberties of information act forms
Game is begin	Game is beginning
Choose right path	Choosing the right path
Problem is examine	Problem is examined
English query	English queries

We then performed Google test for each pair of query set (table 2) and precision values are computed as shown below in the Tables 3 & 4.

TABLE 3. Precision Computation For Queries With Root Words On Google (Using Table 2)

Query	Doc. Retrieved	Precision @10
1.1	5,350,000	0.55
2.1	35,100,000	0.57
3.1	71,800,000	0.5
4.1	114,000,000	0.66
5.1	25,500,000	0.66
6.1	68,900,000	0.77
7.1	125,000,000	0.77
8.1	5,990,000	0.37
9.1	17,800,000	0.88
10.1	2,900,000,000	0.66
11.1	369,000	0.66
12.1	112,000,000	0.62
13.1	1,270,000	0.88
14.1	126,000,000	0.66
15.1	572,000,000	0.5
16.1	18,700,000	0.6
17.1	17,456,000	0.62
18.1	18,187,000	0.7
19.1	26,432,000	0.57
20.1	9,876,000	0.66
Mean Average Precision = 0.643		

Table 4. Precision Computation for Queries without Root Words on Google (USING TABLE 2)

Query	Doc. Retrieved	Precision @10
1.3	7,920,000	0.55
2.2	26,000,000	0.55
3.2	162,000	0.44
4.3	94,300,000	0.44
5.2	27,200,000	0.57
6.3	68,400,000	0.62
7.3	26,300,000	0.44
8.2	2,780,000	0.77
9.2	7,870,000	0.37

10.2	572,000,000	
11.2	194,000	0.77
12.2	10,200,000	0.44
13.2	1,980,000	0.62
14.2	112,000,000	0.55
15.2	556,000,000	0.55
16.2	15,600,000	0.5
17.2	12,768,000	0.55
18.2	13,145,000	0.6
19.2	23,564,000	0.44
20.2	7,956,000	0.57
Mean Average Precision = 0.5445		

From the Tables 3 & 4, it is clear that when queries are in root form, search engine generally indexes more documents (comparing columns II of tables 3 & 4) i.e. the documents Retrieved are higher. The mean average precision for the root word queries is also higher. It shows that the root word queries are better understood by the Search Engines.

Synonymity

It is the common characteristics of most of the natural languages. A query term can have a number of representations by its synonym. We observed while working on English language search engines that any word can express a myriad of implications, connotations, and attitudes in addition to its basic 'dictionary' meaning. Choosing the right word can be difficult for people. In order to justify this impact of varying synonyms on the web search results, we selected another 20 query set with the help of web query logs. The table 5 below shows the set of queries, where each query been regenerated with a synonyms for one of the terms of query (in bold). The queries of table 5 are examined on the Google search engine and precision is computed for each query it is shown in table 6 & 7.

Table 5: Test Query set for synonymity word analysis

Original query	Query with synonyms
School bus safety	School bus security
Aircraft protection act 2004	Aircraft security act 2004
beds sharing with children	beds sharing with kids
freedoms of information act forms	liberty of information act forms

Aim of project	Objective of project
Top beautiful actress in bollywood	Top gorgeous actress in bollywood
application of internet	uses of internet
Advantage of computer	Merits of Computer
Disadvantage of Computer	Demerits of Computer
Ganga is a large river.	Ganga is a big river.
Atom is made up of tiny particles	Atom is made up of small particles
This is correct answer	This is right answer
Game is start.	Game is begin
This answer is wrong	This answer is false
feel very sleepy	feel very tired
shut the door	close the door
house of rabbit	home of rabbit
difficult problems of algebra	hard problems of algebra
images of caps	images of hats
birthday gift	birthday present

Table 6: Precision Computation for Synonymity Using Google (Using Table 5)

Query	Doc. Retrieved	Precision @10
1.1	57,200,000	0.44
2.1	2,090,000	0.77
3.1	4,180,000	0.62
4.1	66,100,000	0.66
5.1	671,000,000	0.66
6.1	5,980,000	0.55
7.1	1,300,000,000	0.5
8.1	11,400,000	0.66
9.1	12,345,000	0.88
10.1	622,800,000	0.77
11.1	564,000,000	0.44
12.1	145,000,000	0.75
13.1	786,000,000	0.66
14.1	111,498,000	0.77
15.1	15,700,000	0.66
16.1	96,000,000	0.44
17.1	14,567,000	0.55
18.1	25,453,000	0.77
19.1	45,600,000	0.77
20.1	15,675,000	0.55
Mean Average Precision = 0.6435		

Table 7: Precision Computation for Synonymy Using Google (Using Table 5)

Query	Doc. Retrieved	Precision @10
1.2	53,800,000	0.66
2.2	40,600,000	0.44
3.2	3,810,000	0.75
4.2	8,040,000	0.44
5.4	662,000,000	0.87
6.2	1,150,000	0.75
7.2	572,000,000	0.55
8.2	98,000,000	0.37
9.2	12,234,000	0.62
10.2	655,700,000	0.55
11.2	675,830,000	0.66
12.2	123,112,000	0.62
13.2	657,000,000	0.62
14.2	104,781,000	0.55
15.2	18,654,000	0.77
16.2	87,678,000	0.33
17.2	12,124,000	0.44
18.2	23,675,000	0.55
19.2	44,134,000	0.66
20.2	16,786000	0.66
Mean Average Precision = 0.593		

The comparative results of the two tables (Table 6 & 7) clearly indicate that search engine (Google) did not properly understand the ‘synonym’ of a query term. That is why its indexing of documents varies in large number on changing the synonym of a query term. The precision values of the corresponding columns (for one query) of two tables also show variations. This would certainly have an impact on search engine’s performance. Our results, however, do not show any trend as to which particular synonym of a query may retrieve more documents and/or higher relevancy.

Sense Ambiguity (Ambiguous Keywords):

Many words are polysemous in nature that is they have multiple possible meaning and senses. Finding the correct sense of the words in the given context is an intricate task. Various researchers (especially Eric Brill [19] and Argaw [20], Navigili and Christopher Stoke [21] and John Tait [22]) have justified the role of Word Sense Disambiguation in the improvement of performance of web searching for English and other languages.

Ambiguous keywords deflate the relevancy of the results. We considered 20 queries (based on our discussion in para III) which are normally ambiguous in nature (a query has been considered ambiguous if one of the term of query is ambiguous). Further, in order to analyze the impact of ambiguity over search engine’s performance we have tried to manually disambiguate each query with the help of Word Net Database and the search engine in consideration and have shown the effect of ambiguity on the performance of the search engines. This is shown in table 8 where the left side column has query with ambiguity and right side column has manually redesigned query without ambiguity same query

Table 8: Test Query Set For Ambiguity Analysis

Query with ambiguous word (in bold)	Query with unambiguous words
Wall paint is blue	Wall color is blue
The train is standing on the platform	The train is standing on the railway platform
There are four seasons in a year	There are four cycle in a year
critical case	critical situation
A bug terminates a program	A error terminates a program
Python are found mostly in rainy Season	Python snakes are found mostly in rainy season
Draw the figure of a flower	Draw the diagram of a flower
Close the door	Shut the door
There should be a break between two lectures	There should be a gap between two lectures
The river is dry	The river is empty
Score of team India in World cup	Run of team India in World cup
balance in my phone	money in my phone
live in present	live in today
aim of a doctor	duty of a doctor
the pitch of sound is high	the level of sound is high
Use of cosine function	Use of cosine expression
The chair of conference	The head of conference
Exercise is necessary to keep our body fit	Physical Exercise is necessary to keep our body fit
interest in science	favorite is science
major accident	big accident

The above queries are examined on the Google search engine and the results are shown below in the Tables 9 & 10.

Table 9: Precision Computation for Ambiguity Using Google (Using Table 8)

Query	Doc. Retrieved	Precision @10
1.1	140,000,000	0.44
2.1	31,600,000	0.66
3.1	2,860,000	0.37
4.1	175,000,000	0.55
5.1	2,550,000	0.5
6.1	1,020,000,000	0.55
7.1	18,400,000	0.66
8.1	435,000,000	0.33
9.1	2,210,000	0.75
10.1	662,000,000	0.37
11.1	4,420,000	0.22
12.1	325,000	0.44
13.1	12,600,000	0.62
14.1	9,260,000,000	0.44
15.1	16,200,000	0.5
16.1	338,000,000	0.55
17.1	174,000,000	0.66
18.1	335,000,000	0.55
19.1	45,100,000	0.44
20.1	683,000,000	0.75
Mean average precision = 0.5175		

Table 10: Precision Computation for Ambiguity Using Google (Using Table 8)

Query	Doc. Retrieved	Precision @10
1.2	374,000,000	0.33
2.2	187,000,000	0.77
3.2	3,150,000	0.44
4.2	374,000,000	0.33
5.2	95,000,000	0.44
6.2	363,000,000	0.55
7.2	66,000,000	0.37

8.2	78,998,000	0.75
9.2	123,000,000	0.44
10.2	112,342,000	0.87
11.2	145,000,000	0.75
12.2	786,000,000	0.66
13.2	111,498,000	0.77
14.2	15,700,000	0.66
15.2	27,200,000	0.57
16.2	68,400,000	0.62
17.2	26,300,000	0.44
18.2	2,780,000	0.77
19.2	572,000,000	0.5
20.2	18,700,000	0.6
Mean average precision = 0.5815		

After examining and comparing the precision values of each queries (Tables 9 & 10), we found that after manual disambiguation of the queries, the precision of 13 out of the 20 queries has improved. The mean average precision has also improved. This shows that the ambiguity in web query can result in poor relevancy of results. Sometimes ambiguity in queries produces adverse results.

2. CONCLUSION AND LIMITATIONS

We have evaluated the performance of the English language search engines in the light of their morphological structures and sense ambiguity.

Our results conclude that the performance of the search engines is quite affected by the morphological issues as well as sense ambiguity problems. Ambiguity is the well known problem of the information retrieval setup. Measures are taken to avoid this problem as it affects the relevancy of the results to a great extent.

In the case of web information retrieval the results of queries vary because web is dynamic in nature. Sometimes the ambiguous query may result out in the relevant results and at another time the similar query may result out in the low relevancy results. Therefore the need of ambiguity detection arises, as automatic disambiguation may lead to the wastage of computational power. Hence detection prior to disambiguation is necessary and it is quite evident from the results.

The sense ambiguity problem much affects the search engine performance because the search engines are not

capable to cope up this problem. Therefore, to resolve this problem there is a need of Word sense disambiguation (WSD) algorithm. This WSD algorithm is used to disambiguate the sense of the ambiguous words and to improve the search engine performance. But before applying the WSD algorithm the ambiguity detection is necessary. It divides the queries in two parts: ambiguous and unambiguous queries. In our thesis, we have design an algorithm to detect the ambiguity in the query. After this the WSD algorithm is applied only on those queries which are ambiguous. This will increase the performance of search engines. By using the WSD methods we develop an algorithm to resolve the ambiguity from the ambiguous queries.

3. ACKNOWLEDGMENT

This paper could not be written to its fullest without the guidance of Dr. Avdesh Gupta, who served as my supervisor, as well as one who challenged and encouraged me throughout my time spent studying under him. He would have never accepted anything less than my best efforts, and for that, I thank him.

REFERENCES

- [1] Sanderson, M., (1994); "Word Sense Disambiguation and Information Retrieval", Proceedings of SIGIR-94, 17th International Conference on Research and Development in Information Retrieval, Dublin, pp. 49-57.
- [2] Krovetz, R; Croft, W. B. "Lexical Ambiguity and Information Retrieval" in ACM Transactions on Information Retrieval Systems, Vol. 10(2), Pp 115 –141, 1992
- [3] Stokoe, C.M. and Jhon, Tait. (2002); "Automated Word Sense disambiguation for Internet Information Retrieval". TREC-2002-WEBTRACK
- [4] Yarowsky, D. "One Sense Per Collocation" In Proceedings of the ARPA Human Language Technology Workshop, Pp 266 – 271, Princeton, NJ, 1993.
- [5] Sanderson, M. "Retrieving with Good Sense" In Information Retrieval, Vol. 2(1), Pp 49 – 69, 2000.
- [6] Bybee, J.L. Morphology: a study of the relation between
- [7] Dwivedi SK, Rastogi P, Goutam R. Impact of language morphologies on Search Engines Performance for Hindi and English language, IJCSIS, Vol. 8, No.3, 2010.
- [8] Voorhees, E.M., & Harman, D. (2001). Overview of TREC 2001. NIST Special Publication 500-250: The 10th text retrieval conference (TREC 2001) (pp. 1-15). Retrieved 17 December 2002 from http://trec.nist.gov/pubs/trec10/papers/overview_10.pdf.
- [9] E. Brill and S. Vassilvitskii, "Using WebGraph Distance for Relevance Feedback in Web Search" in Proceedings of SIGIR'06, Seattle, Washington, USA, pp. -147-153, 2006.
- [10] A. A. Argaw, "Amharic-English Information Retrieval with Pseudo Relevance Feedback", in Proceedings of 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, pp. 119-126, 2007
- [11] R. Navigli and P. Velardi, "Structural Semantic Interconnection: a knowledge-based approach to Word Sense Disambiguation", in Journal of Pattern Analysis and Machine Intelligence, Volume 27, Issue 7, pp. 1075 – 1086, July 2005
- [12] C. Stokoe and J. Tait, "Towards a Sense Based Document Representation for Internet Information Retrieval", in Proceedings of SIGIR'03, July 28- August 1, Toronto, Canada, pp. 791-795, 2003
- [13] Dwivedi SK, Rastogi P. An Entropy based method for removing web query ambiguity in Hindi language, Journal of Computer Science 4(9): 762-767, 2008 ISSN 1549-3636.
- [14] Miguel Angel Rios Gaona, 2009. Web-Base Variant of the Lesk approach to word sense disambiguation. IEEE DOI 10.1109/MICAL.2009.41.
- [15] [34]Roberto Navigli and Giuseppe Crisafulli, 2010. Inducing Word Senses to Improve Web Search Result Clustering. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 116-216, MIT, Massachusetts, USA, 9-11 October 2010.

Optimization of Software using Genetic Algorithm and Formulation of Test Suite for Code Coverage

Mohd Athar¹, Avdhesh Gupta²

¹Research Scholar, Shri Venkateshwara University, Gajraula, India
m.atharjmi@gmail.com

²IMS Engineering College, Ghaziabad (U.P.) India
avvipersonal@gmail.com

Abstract: The objective of programming is not to accomplish a result, but relevance and correctness of the result also required. Correctness can be tested by applying software testing on the developed program. Testing is most important practice which is performed for supporting better quality product. Efficient ways can reduce percentage of cost and time incurred in testing. In spite of scads of theoretical work in field of Software Testing, its advancement is slow towards automation. In this approach, Genetic Algorithm (GA), which is a meta-heuristic algorithm, is employed for optimizing path testing to achieve total code coverage. Random testing is used as a comparison of the efficiency and effectiveness of test data generation using Genetic Algorithms. The advantage of GAs is that through the search and optimization process, test sets are improved such that they are at or close to the input sub domain boundaries. The GAs gives most improvements over random testing when these sub domains are small. Optimization of software testing is achieved by employing GA and the process is automated. It results in formulation of test suite for a module that gives 100 % code coverage. The process of code analysis to find all modules in a program, generation of CFG, finding cyclomatic complexity, determination of all independent paths and GA steps are automated.

IndexTerms: Genetic Algorithm, BBT, SUT.

1. INTRODUCTION

Software testing is important but it possesses some fundamental challenges. It poses two essentially arduous jobs; selecting tests and assessing test results. Selecting test cases are hard as there is enormous number of potential test inputs in varied sequences but only some of them unwrap failures. In Evaluation/assessment, the real output of test run is compared with expected result. This evaluation is done in opaque-testing. Test Suite (TS) generation from operational profile can be automated but it poses substantial hardheaded problems. Time plays a foremost constraint in case of testing. Another vital component is cost. Due to these two constraints, it is intricate task to execute all test cases. When coverage is taken as optimization parameter then target is formulation of TS that could give 100 % code coverage. Optimization problems can be unbridled by GA which can be regarded as computer model of biological evolution. It works on principle of evolution, where superior

chromosomes (having greater fitness value) are chosen for mutation and crossover operations. Evolution continues until the optimized solution is achieved. Good results are found astoundingly speedily when GA is implemented. Generating optimized TS is meta-heuristic problem which can be resolved by GA. Testing tools can be put in two class; dynamic & static.

It's important to have adequate test cases for accomplishment of testing and making software more dependable [2]. Making system reliable is vital as flunking it could sustain massive losses. In [2], BBT is optimized by applying GA. It's implemented in Matlab [version no 7]. Test cases of SUT are heavily influenced by GA. GA depends on various parameters. Population size is vital parameter. Bigger population size brings variation in initial populace at cost of more function evaluations and longer completing times.

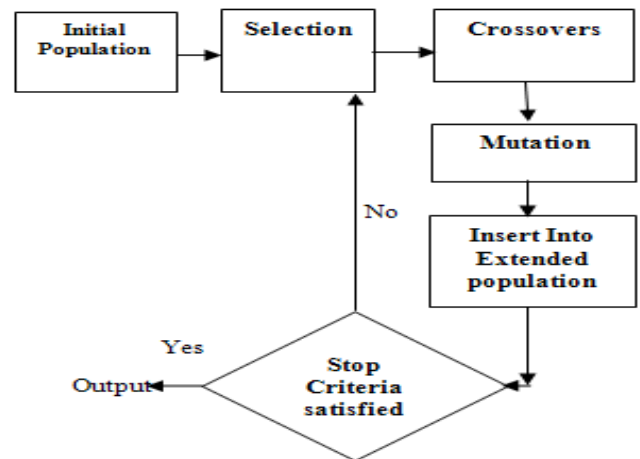


Fig. 1. Proposed solution

Figure 1 shows the procedure adopted in [2]. Fitness function is defined. Threshold fitness is set. Population size is set, crossover rate taken 0.6, mutation rate taken .001. While halt condition is not met, reproduce by crossover and mutation

Testing job is reckoned to be optimization problem whose intent is maximization of noticing errors with minimization of effort. GAs with specification can obtain results with superior quality in lesser time [2]. Figure 6 shows, experiments have been done on “program to find largest number” Inputs are obtained by program’s specification.

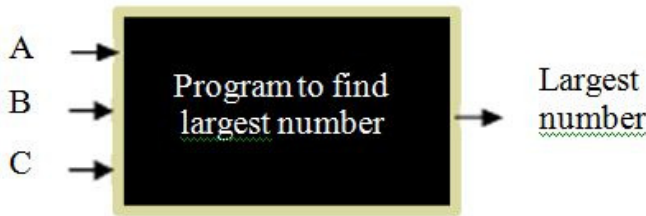


Fig. 2. Block representation of input to the program and output obtained

Figure 2 shows block representation of program designed to generate test cases. The program is to finding largest number among inputted numbers. It is having multiple inputs and one output variable.

Two central issues in process of evolution of genetic search are population diversity and selective pressure. There may be chances of converging early to local optimum solutions if strong pressure is given on selection. On other hand, weaker selection pressure may leads to unproductive search and optimization [3]. Fitness is the key for selection of parents. Fitter chromosomes have more likelihood to get selected. Crossover works on two chromosomes which are chosen as parents to construct two children. In this research work [3], point crossover is employed; two parents are divided partly, with both children getting half of each of parents. Mutation is applied to amend genes in chromosome. For example, let a chromosome be: ‘abdfag’, Mutation may pick gene at position 2 and transform it to a ‘z’, thus, ensuing a new chromosome: ‘azdfag’.

In the work done by Dhawan et.al.[3], test data for input set is delimited in terms of stipulations which illustrate valid and invalid data values. Stipulations are determined from program’s specification.

Role of fitness function to improve WBT

GA is used by WBT to search for precise test data which give high coverage of SUT. Fitness function is necessary feature of GA and is engineered on basis of SUT [10]. Objective function is used to construct fitness function which is applied to sequent genetic ops. Intent of GA is to maximize fitness function. If fitness function is modeled well, probability of reaching higher coverage is enhanced considerably. Based on CFG and requisite test aim, test criteria are separated in different classes [10]:

- **Node-oriented methods:** It requires traversal of particular nodes in CFG. Statement test and condition test can be categorized in this class. Accomplishment of partial aim of this method isn’t reliant on path executed in CFG.
- **Path-oriented methods:** It requires traversal of definite path in CFG. This class comprise of every variation of path test. Finding fitness functions for this class of test is less sophisticated compare to node-oriented method.

Other test criteria can be node-path-oriented method and node-node-oriented method [10]. Baresel et.al [10] separates test into partial objectives and fitness functions are defined for each partial objective, i.e., each statement corresponds partial objective when applying coverage criterion. Ultimate goal of fitness function can be summarized as:

- Substantially enhance chance of detecting solution and attain improved coverage of SUT
- Reduce count of iterations to achieve optimization.

Intention of Baresel et.al [10] is to better formulation of fitness functions, so that, evolutionary testing could be enhanced by getting prominent coverage. It is difficult to investigate reasons behind unsuccessfulness of optimizations because of large search space and existence of many dimensions.

Operators of GA

Execution of GA commence with stochastic population of chromosomes. Fitness function assist evaluation of population and reproductive chances is allowed to population which symbolizes a more adept solution to problem. Chromosomes having superior fitness value are selected by Selection operator. Selection operator is crucial in GA. Jadaan et.al. [12] has proposed altered Roulette Wheel Selection(RWS) to reduce incertitude in selection process. RWS probabilistically choose individuals based on their fitness values (F). Fittest chromosome takes largest share within roulette wheel and chromosome with least fitness value takes smallest share. A random number is generated in interval $[0, S]$ where S is $\sum F$, chromosome whose segment is closer to random number is picked.

2. MOTIVATION

Software testing is a principal technic which is employed for bettering quality attributes of Software Under Test (SUT), particularly reliability and correctness but is also regarded to be tedious. This is also supposed to be intricate work. Software testing suffers from the cognitive biasing of the testers. Automation of testing is a proficient way which can foreshorten time taken and cost incurred in software development. It can also notably better the quality of software.

3. PROBLEM STATEMENT

“To automate generation of TS for each module of SUT by applying GA that could give 100% code coverage”

Random generation of TS does not certify the traversal of all segments of code. There may be odds that the code segment which is not checked could end the program abnormally. So, to obtain optimized TS from set of many feasible TSs, GA is applied..

4. APPROACH

Intent is to optimize TS which could give 100 % code coverage. This optimization which is grounded on total code coverage needs that inner composition of program is well-known. Inner composition of program can be discovered by Path testing in which a set of test-paths are selected in a program. The different independent paths in the program could be determined through CFG. An independent path is that path in CFG that has one novel set of processing statements or novel conditions. Test cases carrying the information of the path covered by them are grouped together to form initial population of chromosomes and GA is applied. In the end, TS is obtained for each module that gives hundred per-cent code coverage.

5. METHODOLOGY

Generating TS that guarantees full coverage of statements in program, is complex task. There are also odds that more than one test case in TS are checking same path. This redundancy is not appreciated. It is imperative to have optimized test data sets. In this section, GA is employed for optimizing Path testing.

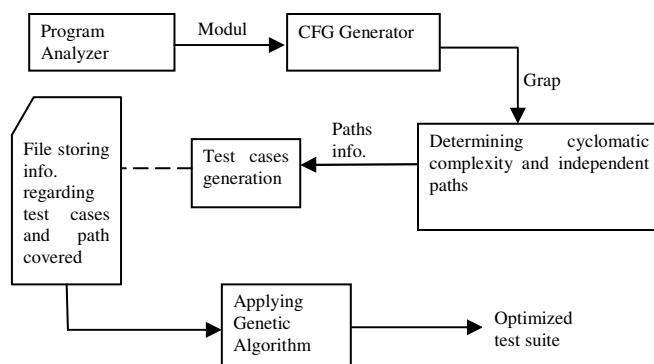


Fig. 3. Block diagram of methodology

Figure 3 illustrates approach applied in this work to accomplish the objective. Program analyzer analyzes the java program and discovers all the modules in it. CFG generator generates the CFG for each module. CFG is used

to find CC and total independent paths. Test cases are generated and paths followed by them are found. The data regarding test cases and path followed are put in a file. This file is utilized when GA is employed. Each of the blocks is explicated fully in this chapter.

Methodology is divided into two approaches:

- Testing
- Applying Genetic Algorithm

6. SURVEY

Generation of chromosomes

Chromosomes are constituted by grouping genes. In GA, chromosomes are optimized. In this work, TS need to be optimized; TS is taken as chromosome, so, test cases are genes. Test cases are randomly grouped to form chromosomes. This grouping form feasible solutions which are optimized to acquire best solution. Initially, size of chromosome is equal to count of total independent paths.

Selection and Crossover operator

Chromosomes having superior fitness value are selected by Selection operator. Reproduction of chromosomes is done by mutation and crossover operator.

Selection operator

This operator is employed to choose parents for mating pool. In this work, roulette selection wheel (RSW) is used as selection operator.

Steps of RSW:

Step 1: Total fitness of all chromosomes in population is calculated.

Step 2: Calculated accumulated frequency for each chromosome.

Step 3: Spawn a random number between zero & total sum of fitness

Step 4: [Loop] Find the sum of accumulated fitness of chromosomes from 0. When sum is greater than random number, return the chromosome.

In this way, parents are selected for mating pool.

Crossover operator

In this work, single point crossover is used.

Steps of single point crossover:

Step 1: Random points are selected in both the parents. This divides the chromosomes into two halves each.

Step 2: Replace second half of first chromosome with second half of second chromosome.

Step 3: Similarly, replace second half of second chromosome with second half of first chromosome.

7. RESULTS AND ANALYSIS

This section discusses regarding implementation of problem statement and goals attempted to accomplish by employing GA. Results got are analyzed graphically. Snapshots have been taken of a sample problem resolved by employing methodology discoursed in chapter 5. It also gives clear picture of implementation.

A sample problem is taken which is a java program. Since WBT is concerned with code structure not functionality, the module is doing simple task of displaying some statements.

```

public class test {
    public static void main(String args[])
    {
        while(true)
        {
            System.out.println("this is while loop");
            break;
        }
        if(1)
            System.out.println("this is cfg test again");
        for(;;)
        {
            System.out.println("this is for loop");
            break;
        }
        for(;;)
        {
            System.out.println("this is for loop again");
            break;
        }
        if(1)
            System.out.println("this is if condition");
        if(1)
            System.out.println("this is if condition again");
    }
}

```

Fig. 4. Sample problem

In first step, program is analyzed to discover the modules in it. Figure 4 is showing the result of code analysis.

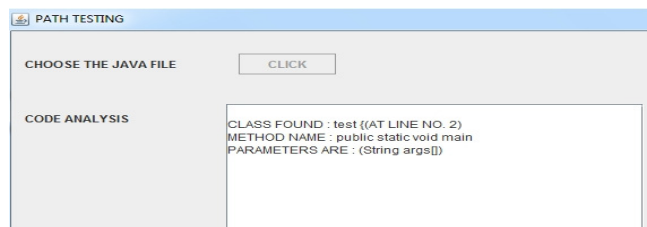


Fig. 5. Output of Code Analyzer in TextArea

As depicted in figure 5, path of java file is given by clicking on “click” button. Class and methods information are displayed in “TextArea”. Information includes “class name”, “methods in class”, “parameters of methods”. *Code Analyzer* also writes output in a “*.txt” file, which is used to fetch line numbers at which method definition exists.

Modules found by *code analyzer* is used by CFG generator to build CFG. *CFG generator* fetches the line number from where module begins from text file generated by *code analyzer*.

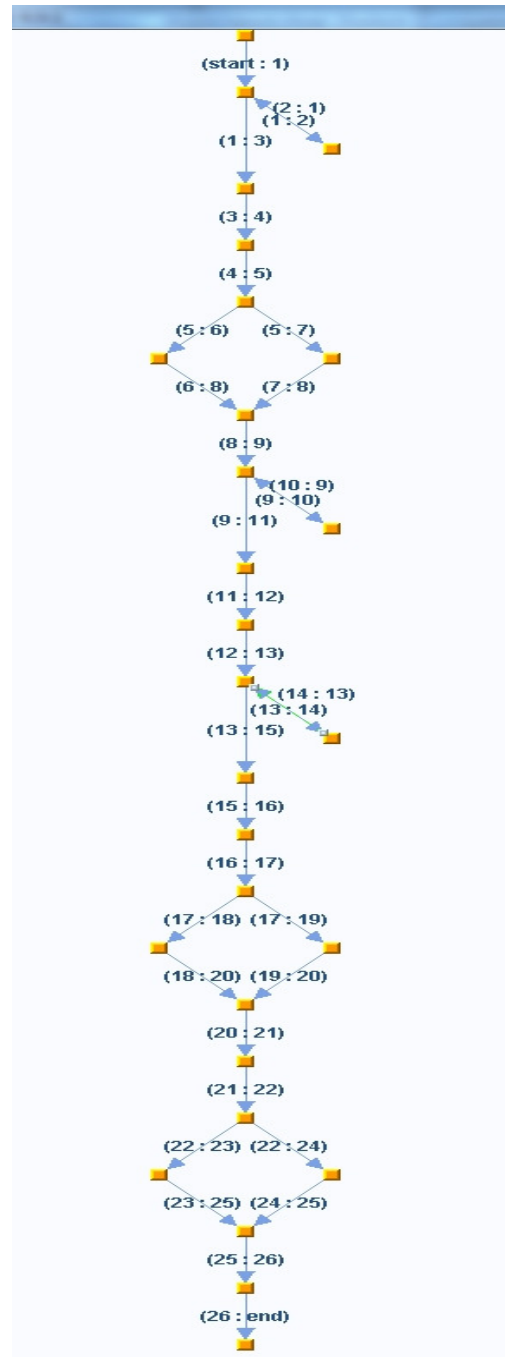


Fig. 6. CFG of main module

Figure 6 shows the CFG of *main* module of sample problem. Orange buttons are vertices of CFG and arrows are edges of CFG. Arrows are labeled like “1:2” showing the flow of control from vertex “1” to vertex “2”.


```

run:
CFG is : {{start, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, end}, {{start,1}, {1,2}, {2,1}, {
Cyclomatic Complexity of module is 11

INDEPENDENT PATHS ARE :
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 19}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 6}, {6 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 19}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 18}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 6}, {6 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 18}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 18}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 6}, {6 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 18}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 19}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 6}, {6 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 : 16}, {16 : 17}, {17 : 18}, {1
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 10}, {10 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 15}, {15 :
{{start : 1}, {1 : 3}, {3 : 4}, {4 : 5}, {5 : 7}, {7 : 8}, {8 : 9}, {9 : 11}, {11 : 12}, {12 : 13}, {13 : 14}, {14 : 13}, {13 : 15}, {15 :

```

Fig. 7. Displaying cyclomatic complexity and independent paths of “main” module

As depicted in Figure 7 cyclomatic complexity of “main” module is 11 and all the independent paths are displayed.

Next step is generation of random test cases. For each test case, corresponding path in CFG is determined. Information regarding the paths covered by test cases is stored in file.

```

run:
OPTIMISING TEST SUITE USING GA
ENTER THE FILE LOCATION TO LOAD TEST CASES
FILE INFO:
test case=tc12 path covered=9
test case=tc13 path covered=10
test case=tc14 path covered=11
test case=tc9 path covered=5
test case=tc8 path covered=6
test case=tc7 path covered=7
test case=tc6 path covered=8
test case=tc5 path covered=3
test case=tc4 path covered=3
test case=tc11 path covered=9
test case=tc3 path covered=1
test case=tc10 path covered=4
test case=tc2 path covered=2
test case=tc1 path covered=2
*****
ENTER THE POPULATION SIZE
30
*****

```

(a) Test cases and path information are loaded from file and initial population is set 30

```

GENERATING INITIAL POPULATION OF CHROMOSOMES
Maximum length of a chromosome can be : 11
INITIAL POPULATION
[tc1, tc10, tc14, tc14, tc4, tc5, tc7, tc7, tc8, tc8, tc9]
[tc1, tc10, tc10, tc13, tc13, tc14, tc2, tc3, tc4, tc6, tc9]
[tc10, tc10, tc11, tc13, tc14, tc14, tc3, tc7, tc8, tc9]
[tc1, tc10, tc14, tc2, tc3, tc4, tc4, tc5, tc7, tc9, tc9]
[tc1, tc1, tc1, tc10, tc11, tc12, tc12, tc2, tc3, tc5, tc9]
[tc1, tc1, tc12, tc12, tc13, tc13, tc2, tc2, tc3, tc5, tc9]
[tc1, tc10, tc10, tc11, tc11, tc12, tc13, tc14, tc4, tc4, tc9]
[tc10, tc10, tc11, tc12, tc13, tc2, tc3, tc3, tc5, tc7, tc7]
[tc11, tc11, tc12, tc12, tc14, tc3, tc5, tc6, tc6, tc7, tc9]
[tc10, tc10, tc13, tc13, tc3, tc4, tc5, tc7, tc8, tc8, tc9]
[tc11, tc11, tc12, tc13, tc14, tc2, tc5, tc5, tc6, tc8, tc8]
[tc1, tc1, tc10, tc11, tc12, tc4, tc4, tc4, tc6, tc6, tc9]
[tc11, tc12, tc12, tc12, tc2, tc3, tc3, tc5, tc5, tc6, tc8]
[tc1, tc12, tc14, tc2, tc4, tc6, tc6, tc7, tc7, tc7]
[tc1, tc11, tc13, tc4, tc4, tc5, tc5, tc5, tc5, tc6, tc9]
[tc10, tc10, tc11, tc13, tc2, tc3, tc4, tc7, tc9, tc9]
[tc1, tc11, tc12, tc2, tc3, tc4, tc5, tc5, tc8, tc8, tc9]
[tc10, tc11, tc2, tc2, tc3, tc3, tc4, tc5, tc5, tc9]
[tc11, tc13, tc3, tc4, tc4, tc6, tc7, tc8, tc8, tc8]
[tc10, tc10, tc12, tc12, tc13, tc13, tc14, tc14, tc4, tc6, tc7]
[tc1, tc10, tc10, tc11, tc11, tc12, tc14, tc4, tc5, tc6, tc8]
[tc10, tc13, tc13, tc14, tc2, tc4, tc4, tc4, tc6, tc6, tc8]
[tc12, tc13, tc2, tc3, tc4, tc4, tc5, tc6, tc7, tc8]
[tc1, tc10, tc10, tc11, tc11, tc12, tc12, tc14, tc5, tc5, tc8]

```

(b) Initial population of chromosomes are generated. Generation 1

```

REMOVING REDUNDANT GENES FROM CHROMOSOMES
[tc1, tc10, tc14, tc4, tc5, tc7, tc8, tc9]
[tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc6, tc9]
[tc10, tc11, tc13, tc14, tc3, tc7, tc8, tc9]
[tc1, tc10, tc14, tc2, tc3, tc4, tc5, tc7, tc9]
[tc1, tc10, tc11, tc12, tc5, tc6, tc7, tc8, tc9]
[tc1, tc10, tc11, tc12, tc2, tc3, tc4]
[tc1, tc12, tc13, tc2, tc3, tc5, tc9]
[tc1, tc10, tc11, tc12, tc13, tc14, tc4, tc9]
[tc10, tc11, tc12, tc13, tc2, tc3, tc5, tc7]
[tc11, tc12, tc14, tc3, tc5, tc6, tc7, tc9]
[tc10, tc13, tc3, tc4, tc5, tc7, tc8, tc9]
[tc11, tc12, tc13, tc14, tc2, tc5, tc6, tc8]
[tc1, tc10, tc11, tc12, tc4, tc6, tc9]
[tc11, tc12, tc2, tc3, tc5, tc6, tc8]
[tc1, tc12, tc14, tc2, tc4, tc6, tc7]
[tc1, tc11, tc13, tc4, tc5, tc6, tc9]
[tc10, tc11, tc13, tc2, tc3, tc4, tc7, tc9]
[tc1, tc11, tc12, tc2, tc3, tc4, tc5, tc8, tc9]
[tc10, tc11, tc2, tc3, tc4, tc5, tc9]
[tc11, tc13, tc3, tc4, tc6, tc7, tc8]
[tc10, tc12, tc13, tc14, tc6, tc7]
[tc1, tc10, tc11, tc12, tc14, tc4, tc5, tc6, tc8]
[tc10, tc13, tc14, tc2, tc4, tc6, tc8]
[tc12, tc13, tc2, tc3, tc4, tc5, tc6, tc7, tc8]
[tc1, tc10, tc11, tc12, tc14, tc5, tc8]
[tc10, tc12, tc13, tc6, tc7, tc8]
[tc1, tc11, tc13, tc14, tc5, tc7, tc9]

```

(c) Redundant genes are removed

```

CALCULATING FITNESS FUNCTION
Fitness of [tc1, tc10, tc14, tc4, tc5, tc7, tc8, tc9] is 0.6363636363636364
Fitness of [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc6, tc9] is 0.7272727272727273
Fitness of [tc10, tc11, tc13, tc14, tc3, tc7, tc8, tc9] is 0.7272727272727273
Fitness of [tc1, tc10, tc14, tc2, tc3, tc4, tc5, tc7, tc9] is 0.6363636363636364
Fitness of [tc1, tc10, tc11, tc12, tc5, tc6, tc7, tc8, tc9] is 0.7272727272727273
Fitness of [tc1, tc10, tc11, tc12, tc2, tc3, tc4] is 0.4545454545454545
Fitness of [tc1, tc12, tc13, tc2, tc3, tc5, tc9] is 0.5454545454545454
Fitness of [tc1, tc10, tc11, tc12, tc13, tc14, tc4, tc9] is 0.6363636363636364
Fitness of [tc10, tc11, tc12, tc13, tc2, tc3, tc5, tc7] is 0.6363636363636364
Fitness of [tc11, tc12, tc14, tc3, tc5, tc6, tc7, tc9] is 0.6363636363636364
Fitness of [tc10, tc13, tc3, tc4, tc5, tc7, tc8, tc9] is 0.6363636363636364
Fitness of [tc11, tc12, tc13, tc14, tc2, tc5, tc6, tc8] is 0.6363636363636364
Fitness of [tc1, tc10, tc11, tc12, tc4, tc6, tc9] is 0.5454545454545454
Fitness of [tc11, tc12, tc2, tc3, tc5, tc6, tc8] is 0.5454545454545454
Fitness of [tc1, tc12, tc14, tc2, tc4, tc6, tc7] is 0.5454545454545454
Fitness of [tc1, tc11, tc13, tc4, tc5, tc6, tc9] is 0.5454545454545454
Fitness of [tc10, tc11, tc13, tc2, tc3, tc4, tc7, tc9] is 0.7272727272727273
Fitness of [tc1, tc11, tc12, tc2, tc3, tc4, tc5, tc8, tc9] is 0.5454545454545454
Fitness of [tc10, tc11, tc2, tc3, tc4, tc5, tc9] is 0.5454545454545454
Fitness of [tc11, tc13, tc3, tc4, tc6, tc7, tc8] is 0.6363636363636364
Fitness of [tc10, tc12, tc13, tc14, tc6, tc7] is 0.5454545454545454
Fitness of [tc1, tc10, tc11, tc12, tc14, tc4, tc5, tc6, tc8] is 0.6363636363636364
Fitness of [tc10, tc13, tc14, tc2, tc4, tc6, tc8] is 0.6363636363636364
Fitness of [tc12, tc13, tc2, tc3, tc4, tc5, tc6, tc7, tc8] is 0.7272727272727273
Fitness of [tc1, tc10, tc11, tc12, tc14, tc5, tc8] is 0.5454545454545454
Fitness of [tc10, tc12, tc13, tc6, tc7, tc8] is 0.5454545454545454

```

(d) Fitness values are calculated

```

BEST CHROMOSOME AT GENERATION 1 is [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc6, tc9] with fitness 0.7272727272727273

```

```

Applying Selection Operator
Parents selected
Fitness of [tc1, tc12, tc13, tc2, tc3, tc5, tc9] is 0.5454545454545454
Fitness of [tc1, tc10, tc11, tc12, tc2, tc3, tc4] is 0.4545454545454545

```

```

Applying Crossover Operator
Children generated
[tc1, tc10, tc11, tc12, tc12, tc13, tc2, tc2, tc3, tc3, tc4, tc5]
[tc1, tc9]
REMOVING REDUNDANT GENES FROM CHILDREN(if any)
Fitness of [tc1, tc10, tc11, tc12, tc13, tc2, tc3, tc4, tc5] is 0.5454545454545454
Fitness of [tc1, tc9] is 0.1818181818181818
Applying Selection Operator
Parents selected
Fitness of [tc10, tc11, tc2, tc3, tc4, tc5, tc9] is 0.5454545454545454
Fitness of [tc10, tc12, tc13, tc14, tc6, tc7] is 0.5454545454545454

```

```

Applying Crossover Operator
Children generated
[tc10, tc11, tc13, tc14, tc2, tc6, tc7]
[tc10, tc12, tc3, tc4, tc5, tc9]
REMOVING REDUNDANT GENES FROM CHILDREN(if any)
Fitness of [tc10, tc11, tc13, tc14, tc2, tc6, tc7] is 0.6363636363636364
Fitness of [tc10, tc12, tc3, tc4, tc5, tc9] is 0.4545454545454545

```

(e) Displaying best chromosome at generation 1. Selection and crossover are applied on population 1

```

Population at generation : 2
Fitness of [tc1, tc10, tc11, tc12, tc13, tc2, tc3, tc4, tc5] is 0.5454545454545454
Fitness of [tc1, tc12, tc13, tc2, tc3, tc5, tc9] is 0.5454545454545454
Fitness of [tc10, tc11, tc13, tc14, tc2, tc6, tc7] is 0.6363636363636364
Fitness of [tc10, tc11, tc2, tc3, tc4, tc5, tc9] is 0.5454545454545454
Fitness of [tc1, tc10, tc11, tc12, tc14, tc3, tc4, tc5, tc6] is 0.6363636363636364
Fitness of [tc10, tc12, tc13, tc14, tc3, tc4, tc5, tc6] is 0.6363636363636364
Fitness of [tc1, tc11, tc12, tc13, tc2, tc3, tc4, tc5, tc6, tc8, tc9] is 0.7272727272727273
Fitness of [tc1, tc11, tc12, tc2, tc3, tc4, tc5, tc8, tc9] is 0.5454545454545454
Fitness of [tc11, tc12, tc13, tc14, tc2, tc5, tc6, tc7] is 0.6363636363636364
Fitness of [tc11, tc12, tc13, tc14, tc2, tc5, tc6, tc8] is 0.6363636363636364
Fitness of [tc10, tc13, tc3, tc4, tc5, tc7, tc8, tc9] is 0.6363636363636364
Fitness of [tc11, tc12, tc2, tc5, tc7, tc8, tc9] is 0.5454545454545454
Fitness of [tc10, tc11, tc12, tc13, tc2, tc3, tc7, tc9] is 0.6363636363636364
Fitness of [tc10, tc11, tc12, tc13, tc2, tc3, tc5, tc7] is 0.6363636363636364
Fitness of [tc11, tc12, tc13, tc14, tc3, tc5, tc6, tc9] is 0.6363636363636364
Fitness of [tc1, tc11, tc13, tc14, tc5, tc7, tc9] is 0.6363636363636364
Fitness of [tc11, tc13, tc14, tc2, tc4, tc7] is 0.6363636363636364
Fitness of [tc10, tc13, tc14, tc2, tc4, tc6, tc8] is 0.6363636363636364
Fitness of [tc1, tc10, tc14, tc4, tc5, tc7, tc8, tc9] is 0.6363636363636364
Fitness of [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc9] is 0.6363636363636364
Fitness of [tc1, tc10, tc14, tc2, tc3, tc4, tc5, tc7, tc9] is 0.6363636363636364
Fitness of [tc1, tc10, tc11, tc12, tc14, tc4, tc5, tc6, tc8] is 0.6363636363636364
Fitness of [tc12, tc13, tc2, tc3, tc4, tc5, tc6, tc7, tc8] is 0.7272727272727273
Fitness of [tc10, tc11, tc13, tc2, tc3, tc4, tc6, tc7, tc8] is 0.8181818181818182
Fitness of [tc10, tc11, tc13, tc2, tc3, tc4, tc7, tc9] is 0.7272727272727273

```

(f) Displaying population at generation 2

```

BEST CHROMOSOME AT DIFFERENT GENERATIONS ARE :
At generation 1, the best chromosome is [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc6, tc9] with fitness value 0.7272727272727273
At generation 2, the best chromosome is [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc7, tc8, tc9] with fitness value 0.8181818181818182
At generation 3, the best chromosome is [tc1, tc10, tc13, tc14, tc2, tc3, tc4, tc7, tc8, tc9] with fitness value 0.8181818181818182
At generation 4, the best chromosome is [tc1, tc10, tc12, tc13, tc14, tc2, tc3, tc5, tc6, tc7, tc8] with fitness value 0.9090909090909091
At generation 5, the best chromosome is [tc10, tc11, tc12, tc13, tc14, tc2, tc3, tc4, tc5, tc6, tc7, tc8, tc9] with fitness value 1.0
BUILD SUCCESSFUL (total time: 3 seconds)

```

(g) Displaying best chromosomes at different generations.

Fig. 8. GA steps to sample problem (a,b,c,d,e,f,g)

Figure 8 shows the implementation of GA on the sample problem. In figure (a) information regarding genes is taken from a file and initial population size is defined. In figure (b) the initial populations of chromosomes are formulated from the genes. In figure (c) redundancy of chromosomes are removed. In figure (d), fitness of chromosomes are displayed. Fitness is calculated by the fitness function as proposed in methodology chapter. In figure (e), working of selection operator and crossover operator are shown. In figure (f), generation 2 is displayed obtained from generation 1 and by applying the GA operators. Figure (g) shows best chromosome of different generations.

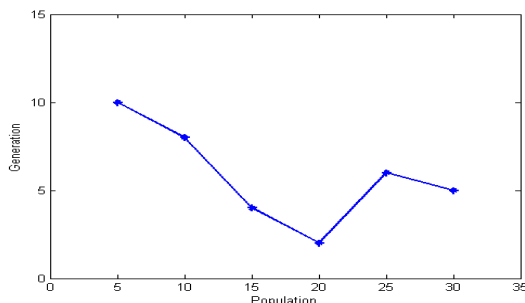


Fig. 9. Initial population vs Generation graph

Figure 9 shows graph of initial population taken and number of generations taken to get an optimized solution. Here initial chromosomes size is 11 and number of test cases provided are 14. The graph shows if population is less, then, GA takes more generations to optimize. But after certain limit, if the population is increased then also GA doesn't perform well.

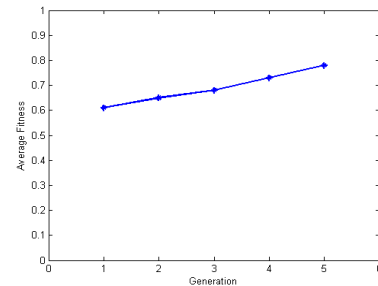


Fig. 10. Generation vs Average fitness graph

Figure 10 shows graph between generation and average fitness of population. Initial chromosome size is 11 and test cases provided are 14 and population size is 30. It is noted that with every passing generation, average fitness of population is improving.

The same methodology is obtained on other problems also.

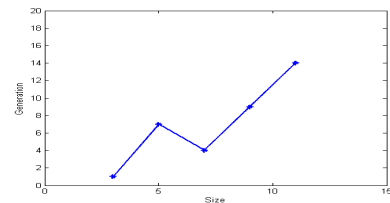


Fig.11. Chromosome size vs. Generation graph

Figure 11 depicts graph between initial size of chromosomes (x-axis) and number of generations (y-axis) taken to get an optimized solution. Initial population is fixed to 5. Chromosome's size is unswervingly proportional to intricacy of module. So, as the intricacy of module increases, it takes more generations to obtain an optimized solution.

8. CONCLUSION AND LIMITATIONS

In this work, optimization of software testing is achieved by employing GA and the process is automated. It results in formulation of test suite for a module that gives 100 % code coverage. The process of code analysis to find all modules in a program, generation of CFG, finding cyclomatic complexity, determination of all independent paths and GA steps are automated. GA is employed on a set of different

software programs and analyses are done on results obtained which decide performance of GA.

In this work, test cases are created manually and paths followed by them are manually determined. RSW selection operator is employed for selecting parents and single point crossover is employed as crossover operator. In future, test case generation from operational profile and path followed by them in CFG can be automated. Other selection operators and crossover operator can be applied and comparison can be drawn between performances of different operators.

REFERENCES

- [1] Bayliss, D. and Taleb-Bendiab, A.: 'A global optimisation technique for concurrent conceptual design', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 179-184, 1994
- [2] BCS SIGIST (British Computer Society, Specialist Interest Group in Software Testing): Glossary of terms used in software testing, 1995
- [3] De Millo R. A. and Offutt A. J.: 'Experimental results from an automatic test case generator', ACM Transactions on Software Engineering and Methodology, Vol. 2, No. 2, pp. 109-127, April 1993
- [4] Feldman, M. B. and Koffman, E. B.: 'ADA, problem solving and program design', Addison-Wesley Publishing Company, 1993
- [5] Frankl P. G. and Weiss S. N.: 'An experimental Comparison of the effectiveness of branch testing and Data Flow Testing', IEEE Transactions on Software Engineering, Vol. 19, No. 8, pp. 774-787, August 1993
- [6] Gallagher M. J. and Narasimhan V. L.: 'A software system for the generation of test data for ADA programs', Micro processing and Microprogramming, Vol. 38, pp. 637-644, 1993
- [7] Gutjahr W.: 'Automatische Testdatengenerierung zur Unterstuetzung des Software tests', Informatik Forschung und Entwicklung, Vol. 8, Part 3, pp. 128-136, 1993
- [8] Hills, W. and Barlow, M. I.: 'The application of simulated annealing within a knowledge-based layout design system', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 122-127, 1994
- [9] Holmes, S. T., Jones, B. F. and Eyres, D. E.: 'An improved strategy for automatic generation of test data', Proc. of Software Quality Management '93, pp. 565-77, 1993
- [10] Jin L., Zhu H. and Hall P.: 'Testing for quality assurance of hypertext applications', Proceedings of the third Int. Conf. on Software Quality Management SQM 95, Vol. 2, pp. 379-390, April 1995
- [11] Korel B.: 'Dynamic method for software test data generation', Software Testing, Verification and Reliability, Vol. 2, pp. 203-213, 1992
- [12] Lucasius C. B. and Kateman G.: 'Understanding and using genetic algorithms; Part 1. Concepts, properties and context', Chemometrics and Intelligent Laboratory Systems, Vol. 19, Part 1, pp. 1-33, 1993
- [13] Müllerburg, M.: 'Systematic stepwise testing: a method for testing large complex systems', Proceedings of the third Int. Conf. on Software Quality Management SQM 95, Vol. 2, pp. 391-402, April 1995
- [14] O'Dare, M. J. and Arslan, T.: 'Generating test patterns for VLSI circuits using a genetic algorithm', Electronics Letters, Vol. 30, No. 10, pp. 778-779, February 1994
- [15] Parmee, I. C. and Denham, M. J.: 'The integration of adaptive search techniques with current engineering design practice', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 1-13, 1994
- [16] Parmee I. C., Denham M. J. and Roberts A.: 'evolutionary engineering design using the Genetic Algorithm', International Conference on Design ICED'93 The Hague 17-19, August 1993
- [17] Rayward-Smith, V. J. and Debus, J. C. W.: 'Generalized adaptive search techniques', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 141-145, 1994
- [18] Reeves, C., Steele, N. and Liu, J.: 'Tabu search and genetic algorithms for filter design', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 117-120, 1994
- [19] Roberts, A. and Wade, G.: 'Optimization of finite wordlength Filters using a genetic algorithm', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 37-43, 1994
- [20] Roper, M.: 'Software testing', International software quality assurance Series, 1994
- [21] Schultz A. C., Grefenstette J. J. and DeJong K. A.: 'Test and evaluation by Genetic Algorithms', U.S. Naval Res. Lab. Washington D.C. USA, IEEE Expert, Vol. 8, Part 5, pp. 9-14, 1993
- [22] Sthamer, H.-H., Jones, B. F. and Eyres, D. E.: 'Generating test data for ADA generic Procedures using Genetic Algorithms', Proc. of ACEDC'94, PEDC, University of Plymouth, UK., pp. 134-140, 1994

A Study on Digital Image Watermarking

Himanshu Goyal¹, Rahul Raj², Arpan Batra³, Kaushlendra Singh⁴

^{1,2,4}M.Tech 1st Year, Dept. of TSE, AITTM, AUUP, Noida

³M.Tech 1st Year, Dept. of ECE, MMEC, MMU, Mullana, Ambala

¹goyalhimanshu19@gmail.com, ²yuvvraaj.rahul@gmail.com,

³arpan.batra1708@gmail.com, ⁴kschahar310@gmail.com

Abstract: In recent years, multimedia authentication techniques have been widely used in the integrity and content authentication of digital media. Current multimedia authentication schemes can be divided into 2 categories according to the authenticator: digital signature-based and digital watermarking-based. This paper reviews several aspects and techniques about digital watermarking. There has been a lot of work conducted in different parts in this field. Watermarking has been used for content protection, copyright management, content authentication and tamper detection. We limit the survey of images only.

IndexTerms: Image and security, Digital Image Watermarking, Types and Attacks

1. INTRODUCTION

Watermarking is not a new phenomenon. For nearly one thousand years, watermarks on paper have been used to identify a particular brand. In the modern era, proving authenticity is becoming increasingly important as more of the world's information is stored as readily transferable bits. Digital watermarking is a process whereby arbitrary information is encoded into an image in such a way that the additional payload is imperceptible to the image observer. Digital watermarking technology is involved with a large number of image processing algorithms and mathematical tools. If we use only the ordinary programming tools to implement the functions of the algorithm and modulation, it is very complex. Therefore, using the high-performance science and engineering calculation software is very important [1].

The difference between digital watermarking and other technology is of three important aspects:

- 1.) Unlike encryption, watermark is imperceptible so that the image will not be detracting from the aesthetic sense.
- 2.) The watermarks and the works they embedded in are inseparable. Even if the works were displayed or converted into other file formats, the watermarks will not be eliminated.
- 3.) The watermarks will have exactly the same transformation experience as the works, which means you can

get the information of transformation by looking at the watermarks.

Today a wide variety of techniques has been proposed but it is quite difficult to classify the approaches and measure their quality. Our intention in this paper is also to discuss the main watermarking parameter as a basis for quality evaluation

The watermarking algorithms discussed in this project are evaluated with the following considerations:

- **Robustness:** A robust watermark will be recoverable in the presence of image manipulation. This includes both unintentional (e.g. noise) and intentional (e.g. cropping, resizing, or compression).
- **Transparency:** A transparent watermark will have little effect on the image quality.
- **Capacity:** The amount of data an algorithm can embed in an image has implications for how the watermark can be applied.
- **Inevitability:** describes the possibility to produce the original data during the watermark retrieval [2].
- **Complexity:** describes the effort and time we need to embed and retrieve a watermark. This consideration is essential if we have real time applications. Another aspect addresses whether the original data in the retrieval process or not. We need to distinguish between non-blind and blind watermarking schemes.

2. TYPES OF WATERMARK

- **Private Watermark:** Private watermarks are also known as secure watermarks. To read or retrieve such a watermark, it is necessary to have the secret key.
- **Public watermark:** Such a watermark can be read or retrieved by anyone using the specialized algorithm. In this sense, public watermarks are not secure. However, public watermarks are useful for carrying IPR information. They are good alternatives to labels.

- *Fragile watermark*: Fragile watermarks are also known as tamper-proof watermarks. Such watermarks are destroyed by data manipulation.
- *Bit-stream watermarking*: The term is sometimes used for watermarking of compressed data such as video.
- *Invisible watermark*: Invisible watermark is hidden in the content. It can be detected by an authorized agency only. Such watermarks are used for content and/or author authentication and for detecting unauthorized copier.
- *Perceptual watermarks*: A perceptual watermark exploits the aspects of human sensory system to provide invisible yet robust watermark. Such watermarks are also known as transparent watermarks that provide extremely high quality contents.
- *Visible watermarks*: Visible watermarks are an extension of the concept of logos. Such watermarks are applicable to images only. These logos are inlaid into the image but they are transparent. Such watermarks cannot be removed by cropping the center part of the image. Further, such watermarks are protected against attacks such as statistical analysis.

The drawbacks of visible watermarks are degrading the quality of image and detection by visual means only. Thus, it is not possible to detect them by dedicated programs or devices. Such watermarks have applications in maps, graphics and software user interface.

3. MODEL OF DIGITAL WATERMARKING

In general, digital watermarking algorithm is composed of two parts: watermark embedding algorithm, the watermark extraction algorithm [3].

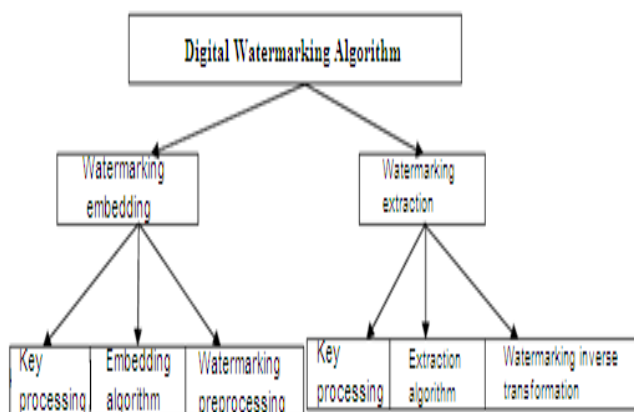


Figure1 : Digital watermarking Algorithm

Fig 2. Shows the general theoretical model of digital watermarking system described below:

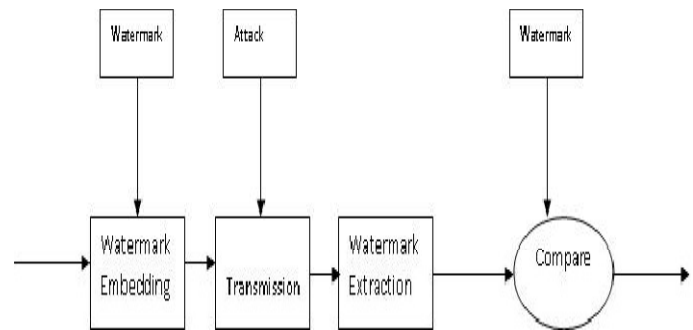


Fig. 2. Process of Watermarking

4. IMAGE WATERMARKING TECHNOLOGIES

Images can be represented as pixels in spatial domain or in terms of frequencies in transform domain. For transferable an image to its frequency representation we use reversible transformation like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), or Discrete Fourier Transform (DFT) [4]. A watermark has been embedded within images by changing these values, i.e. the pixel values or the transform domain coefficients [5].

A. The DCT (Discrete Cosine Transform) Domain Watermarking

Adaptive watermarking techniques have a bit more difficult in the spatial domain. Both the robustness and quality of the watermark should be improved if the properties of the cover image must be exploited. For instance, it is actually preferable to hide watermarking information in noisy regions and edges of images, rather than in smooth regions. The benefit is two-fold; Degradation in smooth regions of an image is more watchable to the HVS, and becomes a prime target for loss compression schemes.

In order for a watermark to be robust to attack, it must be placed in perceptually significant areas of the image. The watermark was based on 1000 random samples of an $N(0, 1)$ distribution. These samples were added to the 1000 largest DCT coefficients of the original image, and the intersect was taken to retrieve the watermarked image. For detection, the watermark was extracted from the DCT of a suspected image. Extraction was based on knowledge of the original signal and the exact frequency locations of the watermark. The correlation coefficient was computed and set to a threshold. If the correlation was large enough, the watermark was detected. Their method was robust to image scaling, JPEG coding, dithering, cropping, and rescanning.

B. The DWT (Discrete Wavelet Transform) Domain Watermarking

The watermark, modeled as Gaussian noise, was added to the middle and high frequency bands of the image. The

decoding process involved taking the DWT of a potentially marked image. Sections of the watermark were extracted and correlated with sections of the original watermark. If the cross-correlation was above a threshold, then the watermark was detected. Otherwise, the image was decomposed into finer and finer bands until the entire, extracted watermark was correlated with the entire, original watermark. This technique proved to be more robust than the DCT method [6] when embedded zero-tree wavelet compression and half toning were performed on the watermarked images.

Many advantages over the wavelet transform, one is that it is believed to more accurately model aspects of the HVS as compared to the FFT or DCT. This allows us to use higher energy watermarks in theoretical regions that the HVS is known to be less sensitive to, such as the high resolution detail bands. Embedded watermarks in these regions allow us to increase the robustness of our watermark, at little to no extra impact on image quality [7].

C. DFT (Discrete Fourier Transform) Domain Watermarking

There are two different kinds of DFT based watermark embedding techniques. One in which watermark is directly embedded and another one is template based embedding. **DFT** is superior for shifting attacks. Shifting in the space domain leads to a phase shift in the frequency domain. In direct embedding watermark is embedded by modifying the phase information within the DFT. A template is a structure which is embedded in the DFT domain to estimate the transformation factor. DFT domain has been explored by researchers because it offers robustness against geometric attacks like rotation, scaling, cropping, translation etc.

5. ATTACKS ON WATERMARKS

In the field of digital watermark, there are various categorizations of attacks on watermarks. These can be categorized by Alit Kulkarni [8] as follows



Figure 3: Scenario Of Changing a Critical Information in an Image

A. Additive Attack:

An adversary or malicious user can augment host by inserting his own watermark W (or several such marks). An

effective additive attack is one in which adversary's mark completely overrides original mark, so that it can no longer be extracted or it is impossible to detect that the original mark temporally precedes the adversary's mark.

B. Subtractive Attack:

In this attack the adversary or malicious user tries to detect the presence, location of the watermark and tries to extract it from the host. An effective subtractive attack is one where the cropped object has retained enough original content to still be of value.

C. Cropping:

This is a very common attack since in many cases the attacker is interested in a small portion of the watermarked object, such as parts of a certain picture or frames of a video sequence. With this in mind, in order to survive, the watermark needs to be spread over the dimensions where this attack takes place.

D. Compression:

This is generally an unintentional attack which appears very often in multimedia applications. Practically all the audio, video and images that are currently being distributed via Internet have been compressed. If the watermark is required to resist different levels of compression, it is usually advisable to perform the watermark insertion task in the same domain where the compression takes place. For instance, DCT domain image watermarking is more robust to JPEG compression than spatial domain watermarking.

E. Distortive Attack:

If an adversary or malicious user has applied some distortive transformation uniformly over the object in order to affect the watermark so that it becomes undetectable/unreadable. An effective distortive attack is one where one can no longer detect the degraded watermark, but the degraded object still has value to the adversary.

F. Multiple Watermarking:

An attacker may watermark an already watermarked object and later make claims of ownership. The easiest solution is to timestamp the hidden information by a certification authority.

G. Filtering:

Low-pass filtering, for instance, does not introduce considerable degradation in watermarked images, videos or audio, but can dramatically affect the performance, since spread-spectrum-like watermarks have non negligible high-frequency spectral contents.

H. Statistical Averaging:

An attacker may try to estimate the watermark and then ‘un-watermark’ the object by subtracting the estimate. This is dangerous if the watermark does not depend substantially on the data. Note that with different watermarked objects it would be possible to improve the estimate by simple averaging. This is a good reason for using perceptual masks to create the watermark.

I. Rotation and Scaling:

It has been very successful with still images. Correlation based detection and extraction fail when rotation or scaling is performed on the watermarked image because the embedded watermark and the locally generated version do not share the same spatial pattern anymore. Obviously, it would be possible to do exhaustive search on different rotation angles and scaling factors until a correlation peak is found, but this is prohibitively complex.

J. Attacks at Other Levels:

There are a number of attacks that are directed to the way the watermark is manipulated. For instance, it is possible to circumvent copy control mechanisms discussed below by super scrambling data so that the watermark is lost or to deceive web crawlers searching for certain watermarks by creating a presentation layer that alters the way data are ordered. The latter is sometimes called ‘mosaic attack’.

6. APPLICATIONS

Many potential applications exist for digital watermarking.

A. Ownership Rights:

Artists and photographers could mark their images to secure ownership rights.

B. Prevent Unauthorized Distribution:

Commercial companies distribute their images to watermark them. Digimarc Corporation already has a software package that searches the Internet for web pages containing specific watermarks.

C. Audio and Video Watermarking:

Watermarking could also apply to other multimedia data

such as audio and video. Compact disks and digital video disks are extremely susceptible to bootlegging via the internet.

7. CONCLUSION

Digital Image Watermarking is a new and merging area of research. A large variety of watermarking techniques is currently available in the literature. It mainly deals with adding hidden messages or copyright notices in digital image. These watermarks, however, are not perfect and more could be done to improve a watermark’s robustness or accuracy in detection. This paper reviews various techniques for image watermarking and attacks on watermarks. As a result, image watermarking is a potential approach for protection of ownership rights on digital image.

REFERENCES

- [1] X. M. Niu, M. Schmucker, C. Busch. Video Watermarking Resisting to Rotation, Scaling and Translation. Proceedings of the SPIE Security and Watermarking of Multimedia Contents IV, San Jose, CA, USA.
- [2] J. Dittmann, P. Wohlmacher, K. Nahrstedt, "Approaches to Multimedia and Security: Cryptographic and watermarking algorithms", unpublished, Villacherstrasse, USA.
- [3] Y. Zhang, "Digital Watermarking Technology: A Review", 2009 ETP International Conference on Future Computer and Communication, Jiangsu, China.
- [4] Potdar V. M., Han S., Chang E.; "A survey of digital image watermarking techniques", Industrial Informatics, 2005. INDIN '05. 2005 3rd IEEE International Conference, Page(s):709 – 716.
- [5] Guan-Ming Su, "An Overview of Transparent and Robust Digital Image Watermarking". Available online at www.watermarkingworld.org/LWMMLArchive/0504/pdf00000.pdf.
- [6] P. Meerwald, A. Uhl, "A Survey of Wavelet-Domain Watermarking Algorithms" EI San Jose, CA, USA.
- [7] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673-1687.
- [8] S. Wansong, M., Zhu, B., Chau, B. and Tewfik, A. (1997) Object-Based Transparent Video Watermarking, Proceedings IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing, Princeton, New Jersey, USA.

Digital Image Watermarking Algorithm Based on DCT and Spread Spectrum

Harsh Vikram Singh

*Department of Electronics Engineering
Kamla Nehru Institute of Technology, Sultanpur (228118)
harshvikram@gmail.com*

Abstract: In this paper, an algorithm for embedding watermarks into host image based on Discrete Cosine Transform and Spread Spectrum has proposed. The proposed algorithm works by dividing the cover into blocks of equal sizes and then embeds the watermark in middle band of DCT coefficient of cover image. Performance evaluation of proposed algorithm has been made using Bit Error Rate (BER) and Peak Signal to Noise Ratio (PSNR) value for different watermark size and images: Lena, Girl, and Tank images yield similar results. This algorithm is simple and does not require the original cover image for watermark recovery. A set of systematic experiments, including JPEG compression, Gaussian filtering and addition of noise are performed to prove robustness of our algorithm.

Keywords: Data embedding, Watermarking, Robust Steganography, Spread Spectrum.

1. INTRODUCTION

The growth of high speed computer networks and that of internet have made reproduction and distribution of digital data easier than ever before. It raises problem of copyright protection. One way for copyright protection is digital watermarking [1-2] which means embedding of certain specific information about the copyright holder (company logos, ownership identification, etc.) into the media to be protected. Digital watermarking is a kind of data hiding technology. It has been used for a variety of applications, including copyright protection, data hiding, and authentication and fingerprinting. Watermarking is a young field and it is growing exponentially [4-5]. Digital watermarking schemes can be categorized as “visible” and “invisible” watermarking. The visible watermarks are easily identified; they are usually not robust against common image processing operation.

The invisible watermarks are more secure and robust than visible watermarks. In invisible watermarking, the watermarked image should look similar to the original image. Based on the processing domain the watermarking schemes can be classified as spatial domain and transform domain [6,7] techniques. The spatial domain watermarking is computationally simple and straight forward wherein host media data is directly replaced by watermark data using

substitution techniques. However, these techniques are more fragile to external attacks and thus provide poor robustness of the watermark. On the other hand the transform domain techniques require more computations but they achieve superior robustness against lossy compressions and different filtering operations such as median, high-pass and low-pass, addition of noise etc[8,9]. Therefore transform domain techniques have proved better choice for achieving enhanced security of watermark and thus for greater assurance of originality of a multimedia data at the receiving end. Generally digital document distribution may consist of image, audio or text or their permutations distributed through open channel. In this paper, our present study focuses on copyright protection on still image documents. Common transform domain techniques mainly are Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) [10, 11], Discrete Fourier Transform (DFT), but DCT is frequently preferred because it is widely used in JPEG and MPEG [12, 13]; and thus it has merited our attention under the present study.

This paper describes an algorithm for achieving enhanced robustness of watermark data. Higher robustness of watermark has been achieved using spread spectrum technique. Embedding of watermark data into mid-band DCT coefficients has been carried out to achieve visual imperceptibility [14, 15] of the hidden watermark which is statistically undetectable and robust against image manipulation attacks. The benefits of the developed algorithm are illustrated through simulation studies by hiding binary logo image into different IEEE standard images such as Lena, Girl and Tank images. The qualitative performance analysis of the suggested algorithm has been carried out through analysis of histogram, JPEG compression, low pass filtering and addition of noise steganalysis techniques [16]. The rest of paper is organized as follows. Section II describes Discrete Cosine Transform and algorithm principle of watermark. Details of the proposed algorithm are presented in Section III. Experiment and results are discussed in Section IV. The performance under various attacks is examined in Section V. Finally, the conclusions have been made in Section VI.

2. MODEL OF ALGORITHM

This section, describes algorithm principle of watermark, Discrete Cosine transform to obtain watermarked image. Watermarked image is combination of cover image and hidden image. DCT is used to convert watermarked object in spatial domain into watermarked image in frequency domain. All the images are assumed to be in standard format.

A. Discrete Cosine Transform

The Discrete Cosine transform has been widely used for source coding in context of JPEG and MPEG and was later also considered for the use of embedding a message inside images and video. It processes some other characteristics and advantages such as vector base good embodiment about image information, small computational complexity, high compression ratio, low error rate, good concealing, and so on, so it is considered the optimal transformation in the digital image processing [18].

Two-dimensional discrete cosine transformation and its inverse transform are defined as [19]:

$$C(u,v) = \alpha(u) \alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

$$f(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u) \alpha(v) c(u,v) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

where, $u, v = 0, 1, 2, \dots, N-1$ $x, y = 0, 1, 2, \dots, N-1$

$\alpha(u)$ is defined as follows: $\alpha(u) = \sqrt{\frac{1}{N}}$ $u=0$; and $\alpha(u) = \sqrt{\frac{2}{N}}$ $u=1, 2, \dots, N-1$

The major benefits of DCT include its high energy compaction properties and availability of fast algorithms for the computation of transform. The energy compaction property of the DCT results in transform coefficients with only few coefficients having values, thus making it well suited for watermarking. Embedding rules in DCT domain are more robust to JPEG/MPEG.

B. Algorithm Principle

The basic principle of digital watermarking algorithm consists of two parts: watermark embedding and recovery. In watermark embedding, original image is first divided into 8x8 sub blocks and then embed watermark bit is spread over middle frequency band DCT coefficient values in the image blocks. The spreading is done by two pseudo-random (PN) sequences, one for zero bit and other for one bit of watermark. At last, watermarked image comes from taking Inverse Discrete Cosine Transform (IDCT). In watermark recovery it is an inverse process of embedding which finds correlation between middle frequency band DCT coefficient

values in the image and corresponding two PN sequences and recovery of watermark.

The middle frequency coefficients are usually chosen due to their moderate variance property. The mid-frequency region is a popular choice for data embedding in order to limit the distortion and enable the algorithm to be robust against a multitude of image manipulating attacks. The mid-frequency regions of the DCT coefficient blocks are used to embed the hidden data as shown in Figure-1, where f_L, f_M and f_H represent the low, medium and high frequency bands respectively [17].

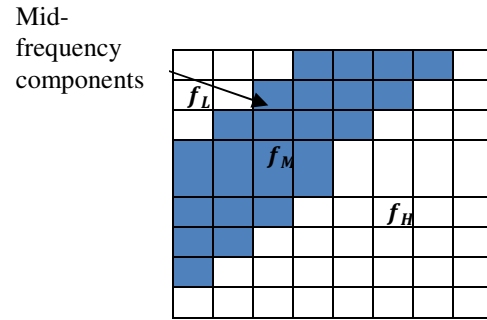


Fig. 1: Mid-frequency DCT region

3. PROPOSED ALGORITHM

The proposed algorithm relies on two pseudo-random (PN) sequences [20], one for zero bit and other for one bit of watermark with low correlation. A pseudo-random noise sequence is generated by using the rand function in MATLAB. This uniform pseudo-random sequence generator must be initialized using a predefined 'key'. This key is available at both embedding and recovery locations and it can be communicated through secure channel prior to sending watermark image over open channel. The proposed algorithm is a combination of spread spectrum watermarking and transform domain watermarking techniques. Use of spread spectrum entails robustness and its combination with DCT domain increases the robustness of this algorithm. The proposed algorithm must provide robustness against a variety of image manipulation attacks.

A. Watermark Embedding

This paper presents an algorithm of digital watermark embedding in the middle frequency band. Instead of using n PN sequences as in [3] here only two PN sequences are used. Fig 2. Illustrates the watermark embedding process. The algorithmic steps are discussed below:

1. Read Cover image and n -bit watermark signal.
2. Generate two PN sequences of length 22 (for 22 mid-band DCT coefficients) using a secret key to reset the

random number generator, one for 'zero' and other for 'one' bit.

3. Transform the original image using 8x8 block 2D-DCT.
4. Hide the i^{th} watermark bit, modulate the i^{th} DCT block of the host using Equation-1 for a '0' or a '1' bit. For $Mi=1$ to n

$$I_W(u, v) = \begin{cases} I(u, v) + \alpha * W_i(u, v), & \text{if } u, v \in f_M \\ I(u, v), & \text{if } u, v \notin f_M \end{cases} \quad (1)$$

Where, f_M Are the mid-band coefficient. α is the gain factor (in present simulation $\alpha = 9$) used to specify the strength of the embedded data; W_i is the appropriate pseudo random noise sequence, based on the i^{th} watermark bit; $I(u, v)$ represents the 8x8 DCT block of the original image $I_W(u, v)$ represents the corresponding watermarked DCT block.

5. To take Inverse transform each of the Watermarked DCT blocks, $I_W(u, v)$, using 8x8 blocks inverse 2D-DCT to get the final Watermarked image $I_W(x, y)$.

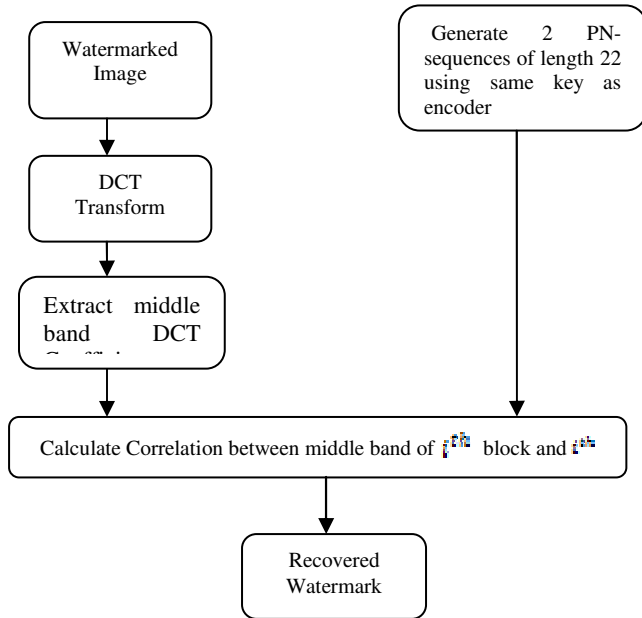


Fig. 2. Watermark Embedding

B. Watermark Recovery

The watermark recovery procedure is based on correlation between the middle frequency band DCT coefficients of the image and corresponding PN sequences [3]. Watermark recovery is the inverse process of the embedding. The steps involved in recovery are listed below:

1. Read watermarked image.
2. Generate two pseudo-random (PN) sequences of length 22 (for 22 mid-band DCT coefficients) after resetting the random number generator using the same secret key as the encoder one for '0' and other for '1'.
3. Transform the watermarked image using 8x8 block 2D-DCT.
4. Extract the middle band coefficients which have recorded the location, determine the watermark information.
5. For $Mi=1$ to n Calculate the correlation between the mid-band coefficients of i^{th} block and i^{th} PN-sequences.
6. Extract the j^{th} watermark bit, b_j , using the following expression

$$b_j = \begin{cases} 0, & \text{if } \text{corr}(0) > \text{corr}(1) \\ 1, & \text{if } \text{corr}(1) > \text{corr}(0) \end{cases} \quad (2)$$

Where $\text{corr}(0)$ is the correlation between extracted coefficient of j^{th} block and PN sequence generated for bit '0'. $\text{corr}(1)$ is the correlation between extracted coefficient of j^{th} block and PN sequence generated for bit '1'.

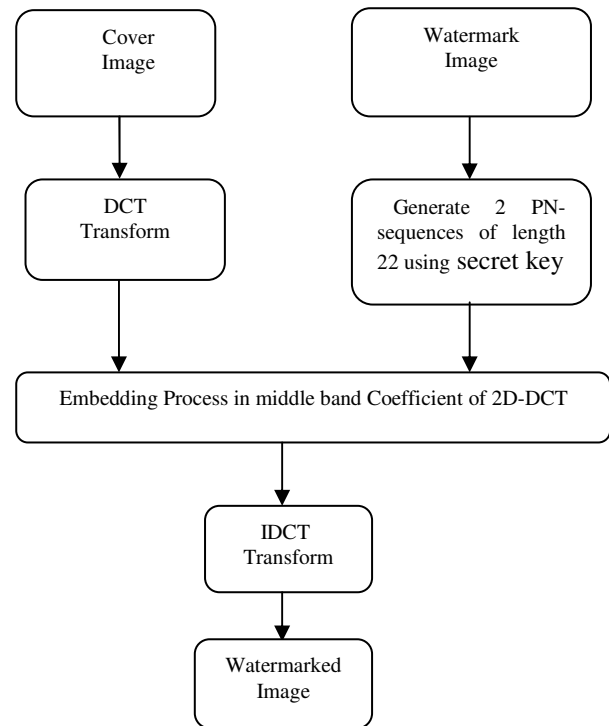


Fig. 3. Watermark Recovery

4. EXPERIMENTS AND RESULTS

The proposed watermarking algorithm is tested with the 512x512 gray Scale Lena image, tests with other images yield similar results. The watermark as shown in figure 4 is used in simulation. The Watermark is binary logo of size (55x52) which is converted into a row vector of size 2860x1 as the watermarking signal, these watermark bits are embedded into the middle band DCT Coefficient of cover image. Performance metrics of watermarking algorithm such as PSNR, BER are computed with and without attacks. The PSNR of watermarking algorithm is reasonable high and the artifacts introduced by watermark embedding are almost invisible. Experimental results without attacks i.e. original image, watermarked image, original and recovered watermarks are shown in Table 1, Figure 4 and 5. Table 2 shows PSNR value for different watermark sizes and images.

Table 1. Performance metrics of watermarking algorithm without attacks

Peak Signal to Noise Ratio (PSNR)	37.29
Bit Error Rate (BER)	0.0178
Processing Time for Embedding	12.95 Sec
Processing Time for Extraction	15.14 Sec



Original Watermark



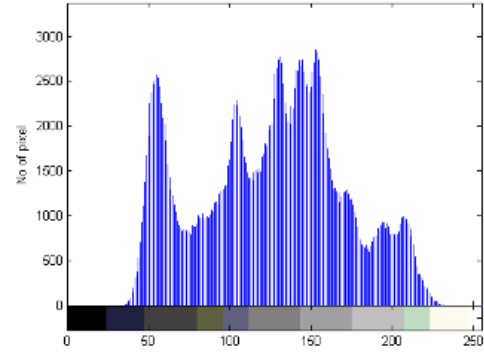
Recovered Watermark

Fig. 4. Original and Recovered watermarks without attacks

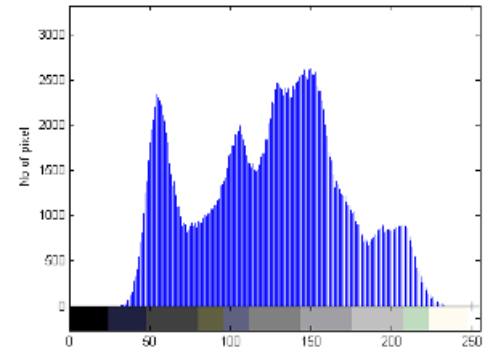


512 x 512 original image 512x512 Watermarked image

Fig. 5. Original and watermarked image without attacks



(a) Original image



(b) Watermarked image

Fig. 6. Histogram of (a) Original image and (b) Watermarked image

Figure 6 shows that there are few pixels having different intensity level of watermarked image as compared to original image due to insertion of watermark.

Table 2. PSNR value for different watermark sizes and images

Images	55x52	64x61	64x17	96x27
Lena	37.2962	37.0212	37.8894	37.7054
Girl	37.2961	37.0212	37.8894	37.7052
Tank	37.2960	37.0212	37.8894	37.7052

5. PERFORMANCE UNDER VARIOUS ATTACKS

A. JPEG Compression: JPEG is important standard for still image compression, so compressed watermarked image at various levels of quality factors and BER is calculated by subjecting the watermarked image to JPEG compression with quality factor 100,80 and 50 to test robustness of the proposed algorithm shown in Table 3.

Table 3: BER value under JPEG Compression with different Quality factor

Quality factor for JPEG compression	Lena	Girl	Tank
	BER	BER	BER
(Q-100)	0.0371	0.0381	0.0357
(Q-80)	0.0290	0.0248	0.0259
(Q-50)	0.0213	0.0161	0.0171

B. Low-pass Filtering: Gaussian low-pass filtering used as another kind of attack. The obtained result in Table 4 shown BER is increasing as gaussian variance (σ) increases.

Table 4: BER of retrieval watermark for a low pass filtering with different gaussian variance (σ)

Variance (σ)	Lena	Girl	Tank
	BER	BER	BER
0.0004	0.0175	0.0171	0.0199
0.0006	0.0213	0.0196	0.0213
0.0008	0.0231	0.0241	0.0234
0.001	0.0271	0.0294	0.0241
0.002	0.0434	0.0490	0.0545

C. Addition of Noise: Adding Salt Pepper noise and Gaussian noise to the watermarked image and by varying noise density in case of salt Pepper noise and variance in case of Gaussian noise. The obtained results show BER is increasing as noise density increases. In case of Gaussian noise BER is increasing as variance increases. The obtained results are tabulated in Table 5 and 6.

Table 5. Performance metric with addition of Salt Pepper noise attack

Noise Density	Lena	Girl	Tank
	BER	BER	BER
0.002	0.0280	0.0245	0.0266
0.004	0.0392	0.0350	0.0364
0.006	0.0647	0.0612	0.0643
0.01	0.0723	0.0731	0.0720
0.04	0.1850	0.1822	0.1829
0.06	0.2311	0.2252	0.2206
0.1	0.2874	0.2941	0.2822

Table 6. Performance metrics with addition of Gaussian noise attack

Gaussian variance (σ)	Lena	Girl	Tank
	BER	BER	BER
0.5	0.0612	0.0556	0.0654
1.0	0.0923	0.0836	0.1049
1.2	0.1133	0.1133	0.1304
1.5	0.1552	0.1811	0.1969
2.0	0.2885	0.3283	0.3290

6. CONCLUSION

This paper proposed a novel digital image watermarking algorithm based on DCT and Spread Spectrum. This algorithm provides statistical security and robustness against various attack. Experimental result demonstrate that the proposed algorithm is resistant to several image manipulating operations and JPEG compression attack. In the case of JPEG compression attacks, even low quality compression (Q-50) resulted in BER of 0.0213 for IEEE standard Lena image i.e more than 97% of the embedded data recovered without any error. The algorithm induces low distortion in the cover image with a PSNR of more than 37 dB.

7. ACKNOWLEDGEMENT

This work is financially supported by Council of Science & Technology, Uttar Pradesh under Young Scientists Scheme. Authors are very thankful to CST for funding.

REFERENCES

- [1] Berghel H and O'Gorman L, "Protecting Ownership Rights through Digital Watermarking", IEEE Computer Mag., pp.101-103, July 1996
- [2] Barmi M, Bartolini F and Cappellini V, "Copyright protection of digital images by embedded unperceivable mark", Image and Vision Computing Vol.16, pp.897-906, 1998.
- [3] Singh Harsh Vikram, Singh S. P, and Mohan Anand, "A New Algorithm for Enhanced Robustness of Copyright Mark," International Journal of Electrical, Computer, and System Engineering, Vol.2, No.2, pp.121-126, 2008
- [4] Petitcolas F. A. P., Anderson R. J. and Kuhn M. G., "Information hiding – a survey," Proc. IEEE, vol. 87, no. 7, pp. 1062–1078, Jul. 1999.
- [5] Johnson, N and Jajodia, S, Exploring Stenography: seeing the unseen, IEEE Computer, 1998, 58(8), 26-34.
- [6] Bruyndockx, J., Quisquater, J., and Macq, B., "Spatial method for copyright labeling," IEEE Workshop on Image Processing, pp. 456-459, 1995.
- [7] Wolfgang, R.B., and Delp, E.J., "A watermarking technique for digital imagery: Further studies," International Conference on Imaging Science, pp. 279-287, 1997.
- [8] Westfeld A., and Pfitzmann A., "Attacks on steganographic systems," Lecture Notes in Computer Science, pp. 61-75, vol. 1768, Springer-Verlag, 2000.
- [9] Schyndel V.R.G., Tirkel A.Z., and Osborne C.F., "A digital watermark," IEEE International Conference on Image Processing, pp. 86-89, 1994.
- [10] Hsieh M.S., Tseng D.C. and Huang Y.H., "Hiding digital watermarks using multiresolution wavelet transform," IEEE Transactions on Industrial Electronics, pp.875 - 882, vol. 48, Issue 5, 2001.
- [11] Koch E. and Zhao J., "Toward robust and hidden image copyright labeling," Proceedings Workshop Nonlinear Signal and Image Processing, Marmaros, Greece, June 1995.

-
- [12] Podilchuk C. and Zeng W., "Watermarking of the JPEG bitstream," in Proceedings of International Conference on Imaging Science, Systems, and Applications (CISST '97), pp. 253-260, Las Vegas, June 1997.
- [13] Guan Y.L. and Jin J., "An objective comparison between spatial and DCT watermarking schemes for MPEG video," Proceedings International Conference on Information Technology: Coding and Computing, pp.207- 211, 2001.
- [14] Podilchuk C. I., "Digital image watermarking using visual models," Proceedings of Electronic Imaging, vol. 3016, San Jose, CA, 1996.
- [15] Podilchuk C. I. and Zeng W., "Perceptual watermarking of still images," Proceedings of Workshop Multimedia Signal Processing, Princeton, NJ, June 1997.
- [16] Wang, H., and Wang, S., "Cyber warfare: steganography vs. steganalysis," Communications of the ACM, pp. 76-82, vol. 47, 2004.
- [17] Amin P.K., Ning Liu, Subbalakshmi K. P. "Statistical Secure Digital Image Data Hiding", IEEE 7th workshop on Multimedia Signal Processing .pp.1-4, 2005.
- [18] Zhang Yu-jin. Image Processing and Analysis, Tsinghua University Press, Beijing, 1998.
- [19] Strang G., "The Discrete Cosine Transform" SIAM Review, Vol. 41, No. 1, pp.135-147, 1999.
- [20] F. J. Mac William and N. J. A. Sloane, "Pseudorandom Sequences and Arrays", Proc. of the IEEE, Vol. 64, No. 12, pp 1715-1729, Dec 1976.

Koch Shaped Fractal End Coupled Microstrip Bandpass Filter

Gurpreet Kaur Kohli¹, Pravesh Singh²

¹Student of M.Tech ECE, KIET, Ghaziabad, India, gurpreetkohliece@gmail.com

²Department of ECE, KIET, Ghaziabad, India, pravesh_singh01@yahoo.co.in

Abstract: Filters are the key elements of telecommunications and radar systems and are important items in the performance and cost of such system. There are various types of Fractal shaped but in this paper the Koch Fractal shaped is used to design a Microstrip End Coupled Half-wavelength Bandpass Filters is simulate. The proposed filter having an center frequency of 2.02 GHz with the Return Loss is 2.434db and Insertion loss is 18.83. The frequency bands development in microwave filter, plays a major role in many RF or microwave applications.

Index Terms - Fractal shape, Microstrip Bandpass Filter, Half wavelength End Coupled Filter, Wavelength resonators and Centre frequency

1. INTRODUCTION

In several microwave devices there are various types of fractal geometries such as Koch curve, Sierpinsky gasket, and Hilbert curve, etc. have been widely used to design, due to their reduction of resonant frequencies, smaller size and broadband width. In this paper the Koch fractal geometry, which named after the mathematician Helge von Koch, is used. There are two unique properties of fractal shape: Space filling and Self similarity. It can be used to filled a limited area as the order increases and occupies the same area regardless of the order. This is due to the space filling property. By self similarity, a portion of the fractal geometry always looks the same as that of the entire structure [11]. The End Coupled Filters are particularly suitable for printed circuit technology for planar formats as easily implemented and has the advantage of taking less space than a plain transmission line.

The minimum width of gaps, like the minimum width of tracks, is limited by the resolution of the printing technology. To reduce insertion loss in the pass-band, the gaps are usually much smaller than the substrate height to enable tight coupling. The resonator lengths depend on the guide wavelength, coupling reactance and the gap capacitance. This configuration provides relatively narrow bandwidth. Since this structure is large, it is not a much preferred configuration. This type of filter has desirable advantages such as low-cost fabrication and easy integration[8].

2. END COUPLED FILTER

Fig 1.1 shows the end-coupled half-wavelength bandpass filters, which has open-end microstrip resonators of almost half-wavelength ($\lambda/2$) long at the midband frequency f_o of the bandpass filter. The resonators are coupled by means of gap capacitances between the resonator sections. The resonator length θ and the coupling gaps S between successive resonators are important design parameters.



Fig 1.1 : General structure of end coupled microstrip bandpass filter.

This filter operates like the shunt-resonators type and the design equations are [3]:

$$\frac{J_{01}}{Y_0} = \sqrt{\frac{\pi}{2} \frac{FBW}{g_0 g_1}} \quad (1)$$

$$\frac{J_{j,j+1}}{Y_0} = \frac{\pi FBW}{2} \frac{1}{\sqrt{g_j g_{j+1}}} \quad \text{for } j=1 \text{ to } n-1 \quad (2)$$

$$\frac{J_{n,n+1}}{Y_0} = \sqrt{\frac{\pi}{2} \frac{FBW}{g_n g_{n+1}}} \quad (3)$$

where $g_0, g_1 \dots g_n$ are the element of a ladder-type lowpass prototype with a normalized cutoff $\Omega_c = 1$ and FBW is the fractional bandwidth of bandpass filter. The $J_{j,j+1}$ are the characteristic admittances of J-inverters and Y_0 is the characteristic admittance of the microstrip line. Assuming the capacitive gaps act as perfect, series- capacitance discontinuities of susceptance $B_{j,j+1}$ are: [2]

$$\frac{B_{j,j+1}}{Y_0} = \frac{\frac{J_{j,j+1}}{Y_0}}{1 - \left(\frac{J_{j,j+1}}{Y_0}\right)^2} \quad (4)$$

$$\theta_j = \pi - \frac{1}{2} \left[\tan^{-1} \left(\frac{2B_{j-1,j}}{Y_0} \right) + \tan^{-1} \left(\frac{2B_{j,j+1}}{Y_0} \right) \right] \text{ radians} \quad (5)$$

where the $B_{j,j+1}$ and are evaluated at f_0 . The second term on the right-hand side of (1,2,3) indicates the absorption of the negative electrical lengths of the J-Inverters associated with the j th half-wavelength resonator.

Table-1: Parameters of End Coupled Half- wavelength Resonator Bandpass Filter

S.No	Parameters	A	B	C	D
1	g_n	1.5963	1.0967	1.5963	1.0967
2	$Z_0 J_n$	0.3137	0.1187	0.1187	0.3137
3	B_n	6.96×10^{-3}	2.41×10^{-3}	2.41×10^{-3}	6.96×10^{-3}
4	C_n	0.4520pF	0.1564pF	0.1564pF	0.4520pF
5	θ_n	2.72 rad	2.91 rad	2.72 rad	-

The coupling gaps $s_{j,j+1}$ of the microstrip end coupled resonator filter are;

$$C_g^{j,j+1} = \frac{B_{j,j+1}}{\omega_0} \quad (6)$$

where $\omega_0 = 2\pi f_0$ is the angular frequency at the midband. The physical lengths of resonators are given by

$$\ell_j = \frac{\lambda_{g0}}{2\pi} \theta_j - \Delta \ell_j^{e1} - \Delta \ell_j^{e2} \quad (7)$$

The effective lengths can then be found by

$$\Delta \ell_j^{e1} = \frac{\omega_0 C_p^{j-1,j}}{Y_0} \frac{\lambda_{g0}}{2\pi} \quad (8)$$

$$\Delta \ell_j^{e2} = \frac{\omega_0 C_p^{j,j+1}}{Y_0} \frac{\lambda_{g0}}{2\pi} \quad (9)$$

3. FILTER DESIGN

From the parameters given below the Fractal shaped End coupled micro strip band pass filter , with a 0.5dB equal-ripple pass band characteristic 1st order is designed using IE3D with the center frequency of 2.45 GHz having a bandwidth of 10% and equal ripple in the pass-band of 0.5dB, with the FR4 substrate of dielectric constant 4.2 with thickness of 1.58mm for a third order Chebyshev filter.



Fig. 1.2. Proposed Filter Design layout (1st Order)

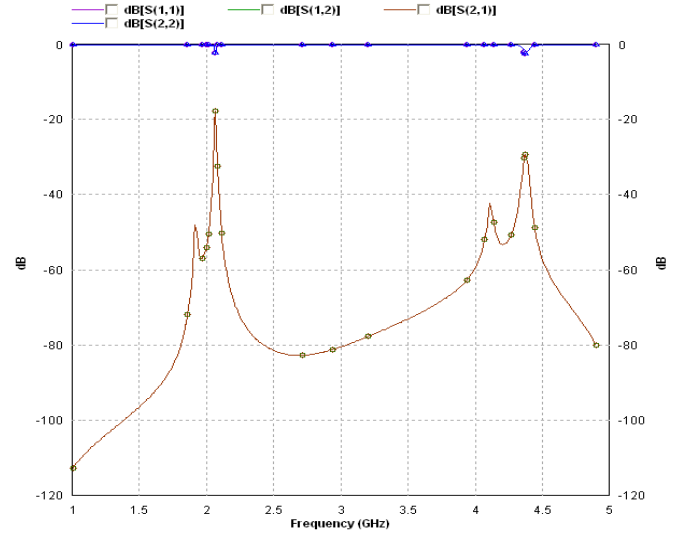


Fig 1.3 : Response for end coupled microstrip band pass filter (1st order)

Practical Design Filters :-



Fig 1.4 . Practical 1st order Fractal shaped Microstrip Bandpass Filter

4. RESULT AND ANALYSIS

It has been seen from the responses using IE3D shows that the end coupled line band-pass filter having an identical near pass-band region for 1st order at 2.02GHz the Return Loss is 2.434db and Insertion loss is 18.83 .

5. CONCLUSION

It is observed from this analysis that the further symmetrical approach i.e. iterations tends to produce a more compact filter with less coupling effect in its realization when frequency increases and also this minimizes required space for realization and is suitable for integration within wireless system.

REFERENCES

- [1] D. M. Pozar, *Microwave Engineering*, New York: John Wiley and Sons, 1998, 2nd Ed., pp. 474-485
- [2] J.-T. Kuo, S.-P. Chen, and M. Jiang, "Parallel-coupled microstrip filters with over-coupled end stages for suppression

- of spurious responses," IEEE Microw. Wireless Compon. Lett., vol. 13, no. 10, pp. 440–442, Oct. 2003.
- [3] Jia-Sheng Hong, M.J Lancaster, "Microstrip Filters for RF/Microwave Applications", Wiley and Sons, 2001.
- [4] J.-T. Kuo and M. Jiang, "Enhanced microstrip filter design with a uniform dielectric overlay for suppressing the second harmonic response," IEEE Microw. Wireless Compon. Lett., vol. 14, no. 9, pp. 419–421, Sep. 2004.
- [5] H. Zhang and K. J. Chen, "A Tri-Section Stepped-Impedance Resonator for Cross-Coupled Bandpass Filters," IEEE Microwave and Wireless Components Letters, vol. 15, no. 6, pp. 401 – 403, June 2005.
- [6] J. -T. Kuo, T. -H. Yeh, and C. -C. Yeh, "Design of microstrip bandpass filters with a dual-passband response," IEEE Trans. Microwave Theory & Tech., vol. 53, no. 4, pp. 1331 – 1337, April 2005.
- [7] C.-M. Tsai, S- Y. Lee, and H.-M. Lee, "Transmission-line filters with capacitively loaded coupled lines," IEEE Transactions on Microwave Theory and Techniques, vol. 51, pp. 1517–1524, 2003.
- [8] S.-M. Wang, C.-H. Chi, M.-Y. Hsieh, and C.-Y. Chang, "Miniaturized spurious passband suppression microstrip filter using meandered parallel coupled lines," IEEE Trans. Microw. Theory Tech., vol. 53, no. 2, pp. 747–753, Feb. 2005.
- [9] R. Saal, E. Ulbrich, "On the design of filters by synthesis", IRE Trans., Vol. CT-5, pp 284–327, 1958.
- [10] U. P. Rooney and L.M. Underkofler, "Printed circuit integration microwave filters", Microwave J., volume 21, pp. 68–73, Sept., 1978.
- [11] IlKwon Kim, Nickolas Kingsley, Matt Morton, John Papapolymerou, Ramanan B., Manos M. Tentzeris and Jong-Gwan Yook, "Fractal-Shaped Microstrip Coupled-Line Bandpass Filters for Suppression of Second Harmonic" IEEE Trans. Microwave Theory Tech., Vol. 53, No. 9, 2005.
- [12] Jawad K. Ali, "A New Miniaturized Fractal Bandpass Filter Based on Dual-Mode Microstrip Square Ring Resonator", 5th International Multi-Conference on Systems, Signals and Devices (2008) .
- [13] Dharendra Kumar and Asok De "Parallel Coupled Microstrip Filter Design using Electromagnetic Bandgap Structure" , International Journal of Contemporary Research in Engg. and Tech. Vol. 1, No. 1, 2011.
- [14] Wen-Ling Chen and Guang-Ming Wang "Effective Design of Novel Compact Fractal-Shaped Microstrip Coupled-Line Bandpass Filters for Suppression of the Second Harmonic" , IEEE Microwave And Wireless Components Letters, Vol. 19, No. 2, February 2009
- [15] Yaqeen S. Mezaal "A New Microstrip Bandpass Filter Design Based on Hilbert Fractal Geometry for Modern Wireless Communication Applications" International Journal of Advancements in Computing Technology Volume 1, Number 2, December 2009.

New Fast and Efficient Progressive Switching Median Filter for Digital Images

Ritesh Kumar¹, Lav Kedia²

¹*School of Electrical Sciences Electrical and Electronics*

²*IIT Bhubaneswar, Bit Mesra*

Abstract: Progressive switching median filter considered the uncorrupted pixels to calculate the median value after some iterations. Here we are proposing the modified progressive switching median filter in which the substitute pixel value is the half of the sum of selected mean and median value. Here we are setting limit also on the selection of good pixel. Experimental results show that this proposed method performs well for high impulse noise ratio.

Keywords: Image Processing, Median Filter, Switching median filter, Progressive switching median filter.

1. INTRODUCTION

This Noise reduction is still the Challenging problem in digital image processing. In digital image noise identification and reduction [1] is an important pre-processing step as images are often corrupted by noise during acquisition and transmission process. In the literature several filters have been studied for noise reduction, in which one of the most famous filter for gray scale image is the median filter [2] where the central pixel intensity is replaced by a median value taken within a filtering window. The median filter is very simple and computationally efficient but it causes the loss of fine image details because it replaces all the pixels of the noisy image by a median value irrespective of them being corrupted or not. The size of neighborhood is determined by squared, sliding window which passes through the whole image pixel to pixel.

There are various filters designed for noise removal that is as per the type of noise present in digital images. This paper is describing the impulse noise removal from digital images while preserving the image details and integrity of the edges simultaneously. Standard median filter [4-9] is the most popular for impulse noise removal but it causes blurring as well as it is not sufficient in presence of high density of impulse noise and showing edge jitters often. Standard median filter is of two types:- (a) Weighted Median Filter(WFM) (b) Switching Median Filter (SMF).

Weighted median filter [7, 9] selectively gives some weight to pixels in the filtering window usually with the central pixel most. Higher weights assigned to the central pixel reduce smoothing effect and makes WMF better in

preserving image sharpness. This paper is going to describe the switching median filter in which no. of pixels are reduced subjected to median filtrations to those that are believed to noisy instead of WMF. Wang and Zhou proposed their progressive switching median filter [4] by implementing Sun and Neuvo's switch I scheme [3] in their impulse detector with a number of progressive iterations. They proposed that the detected noisy pixel is also removed by the progressive iterative manner. We are proposing the new fast and efficient progressive switching median filter in which the detector algorithm is by modified version of those proposed by Wang and Zhou. Here we are substituting the corrupted pixel value by half of the sum of mean and median value. This paper is organized in this way that Section II describes the impulse noise model and section III is giving idea about the progressive switching median filter and modified PSM. Section IV demonstrates the results and discussion and Section V is the conclusion of the proposed filter.

2. IMPULSE NOISE MODEL

The PDF of (bipolar) impulse noise is given by

$$P(z) \begin{cases} P_a \text{ for } z = a \\ P_b \text{ for } z = b \\ 0 \text{ otherwise} \end{cases}$$

If $b > a$, gray level b will appear as a light dot in image. Conversely, level a will appear like a dark dot. If either P_a or P_b is zero, the impulse noise is called unipolar. Impulse noise is found in situations where quick transients, such as faulty switching take place during imaging. Four impulse noise models are reported in recent papers [5].

A. Impulse Noise Model 1:

Noise is modeled as salt-and-pepper impulse noise. Pixels are randomly corrupted by two fixed extreme values, 0 and 255 (for 8-bit monochrome image), generated with the same probability. That is, for each image pixel at location (i, j) with intensity value S_{ij} , the corresponding pixel of the noisy image will be x_{ij} , in which the probability density function of x_{ij} is :

$$f(x) = \begin{cases} \frac{p}{2} \text{ for } x = 0 \\ 1 - p \text{ for } x = S_{i,j} \\ \frac{p}{2} \text{ for } x = 255 \end{cases}$$

Where P is the noise density.

B. Impulse Noise Model 2:

For the Model 2, it is similar to Model 1, except that each pixel might be corrupted by either “pepper” noise (i.e., 0) or “salt” noise with unequal probabilities. That is.

$$f(x) = \begin{cases} p_1 \text{ for } x = 0 \\ 1 - p \text{ for } x = S_{i,j} \\ p_2 \text{ for } x = 255 \end{cases}$$

Where $p = p_1 + p_2$, and $p_1 \neq p_2$.

C. Impulse noise model 3:

Instead of two fixed values, impulse noise could be more realistically modeled by two fixed ranges that appear at both ends with a length of m each, respectively. For example, if m is 10, noise will equal likely be any values in the range of either [0, 9] or [246, 255]. That is :

Where $p = p_1 + p_2$, and $p_1 \neq p_2$.

3. MODIFIED PROGRESSIVE SWITCHING MEDIAN FILTER

Progressive switching median filter detects the corrupted pixels of the noisy image by setting a flag in all the respective positions of the pixels whose absolute difference from the median exceeds a predefined threshold value in an iterative manner. Best restoration result is obtained when total no of iteration is 3 and the threshold value T is taken as 40. The corresponding median value with 3x3 window size for each pixel $x_{(i,j)}$ is :

$$m_{(i,j)} = \text{Med}\{x(i,j) | i-1 \leq i \leq i+1, j-1 \leq j \leq j+1\}$$

Initially, the number of noisy pixels N_I that have been detected is set as : $N_I = 0$. To determine pixel $x_{(i,j)}$ is impulse or not the absolute difference between $m_{(i,j)}$ and $x_{(i,j)}$ is considered, If impulse found N_I is increased by 1.

$$N_I = N_I + 1 \text{ if } |m_{(i,j)} - x_{(i,j)}| \geq T_1$$

The predefined value of T_1 is 40. If N is the total no. of pixels present in the image, the noise ratio RE is given by:

This ratio is determined when all the filtering operation on all pixels have done once. The value of window size W and the threshold value is calculated as :

$$W = \begin{cases} 3, \text{ if } R_E \leq T_R \\ 5, \text{ if } R_E > T_R \end{cases}$$

$$\text{and } T = a + bR_E$$

Where T_R , a and b are defined as 25%, 65 and -50 respectively.

A. Impulse Detection

Two image sequences are produced after impulse detection. One is a sequence of gray scale image and second one is a binary flag image sequence where flag sequence is used to indicate whether the pixel at position (i, j) is impulse or not after n-th iterations. $x^n_{(i,j)}$ represents the pixel value at position (i, j) in the image after n-th iterations and the flag value $f^n_{(i,j)}$ indicates the pixel at position (i, j) is impulse or not as shown :

$$f^n_{(i,j)} = 0, \text{ Good Pixel}$$

$$f^n_{(i,j)} = 1, \text{ Impulsive pixel}$$

$$\text{Initially, } f^0_{(i,j)} = 0$$

The median value after n-th iteration with window size WXW is determined as :

$$m^{(n-1)}_{(i,j)} = \text{Med} \left\{ \begin{array}{l} x^{n-1}_{(i,j)} \mid i - \frac{w-1}{2} \leq i \leq i + \frac{w+1}{2} \\ j - \frac{w-1}{2} \leq j \leq j + \frac{w+1}{2} \end{array} \right\}$$

If the difference between the $x^{n-1}_{(i,j)}$ and $m^{(n-1)}_{(i,j)}$ exceeds the threshold value T, the flag $f^{n-1}_{(i,j)}$ is set to 1.

$$f^n_{(i,j)} = \begin{cases} f^{n-1}_{(i,j)} & \text{if } |x^{n-1}_{(i,j)} - m^{(n-1)}_{(i,j)}| < T \\ 1 & \text{else} \end{cases}$$

And

$$x^n_{(i,j)} = \begin{cases} m^{n-1}_{(i,j)} & \text{if } f^n_{(i,j)} \neq f^{n-1}_{(i,j)} \\ x^{n-1}_{(i,j)} & \text{if } f^n_{(i,j)} = f^{n-1}_{(i,j)} \end{cases}$$

B. Filtering Scheme

Noise filtering scheme also produces two image sequence, one is a sequence of gray scale image and the other is a sequence of binary flag image. Where $y^n_{(i,j)}$ represents the

pixel value at position (i, j) in the image after n -th iteration and $g_{(i,j)}^n$ is used to indicate whether the pixel at position (i, j) is impulse or not after the n -th iterations.

Initially,

$$g_{(i,j)}^0 = f_{(i,j)}^n$$

Only good uncorrupted pixels are chosen to calculate the replaced value. Our method is substituting the corrupted pixels to the half of the sum of its median value $m_{(i,j)}^{n-1}$ and mean value $mn_{(i,j)}^{n-1}$ as defined as :

Where,

$$m_{(i,j)}^{n-1} = \text{Med} \begin{cases} y_{(i,j)}^{n-1} | g_{(i,j)}^{n-1} = 0, \\ |i - \frac{w_f - 1}{2} \leq i \leq i + \frac{w_f - 1}{2} \\ |j - \frac{w_f - 1}{2} \leq j \leq j + \frac{w_f - 1}{2} \end{cases}$$

And

$$mn_{(i,j)}^{n-1} = \text{Mean} \begin{cases} y_{(i,j)}^{n-1} | g_{(i,j)}^{n-1} = 0, \\ |i - \frac{w_f - 1}{2} \leq i \leq i + \frac{w_f - 1}{2} \\ |j - \frac{w_f - 1}{2} \leq j \leq j + \frac{w_f - 1}{2} \end{cases}$$

$y_{(i,j)}^n$ is the modified only when the pixel (i, j) is an impulse and M is at least equal to G_p , where M denotes the total number of good pixels with $g_{(i,j)}^{n-1} = 0$ in the $W \times W$ window, and G_p denotes the number of good pixels that should be used in finding the corresponding median and mean values. $y_{(i,j)}^n$ is defined as :

$$y_{(i,j)}^n = \begin{cases} S_{(i,j)}^{n-1} \text{ if } g_{(i,j)}^{n-1} = 1 : M \geq G_p \\ y_{(i,j)}^{n-1} \text{ else} \end{cases}$$

When the condition is not satisfied the noisy pixel is left unprocessed until further iterations to restore the noisy pixel. The impulse pixel is considered good after modifications in the subsequent iteration is given as :

$$g_{(i,j)}^n = \begin{cases} g_{(i,j)}^{n-1} \text{ if } y_{(i,j)}^n = y_{(i,j)}^{n-1} \\ 0 \text{ if } y_{(i,j)}^n = S_{(i,j)}^{n-1} \end{cases}$$

Noise filtering iteration stops when $\sum g_{(i,j)}^n = 0$ and the restored output $y_{(i,j)}^n$ is obtained.

4. RESULTS AND DISCUSSIONS

The performance of the filter is expressed by two important measures RMSE and PSNR that is defined as:

A. Root Mean Squared Error (RMSE)

Suppose that the original image $x_{(i,j)}^n$ of size $M \times N$ has been denoised, using an image denoising scheme, and $y_{(i,j)}^n$ be the denoised estimate. The RMSE between the denoised image and the original image is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (y_{(i,j)}^n - x_{(i,j)}^0)^2}{M * N}}$$

B. Peak Signal to Noise Ratio (PSNR)

It is inversely proportional to the RMSE, its units are in decibels (dB) and is formally defined by

$$PSNR = 20 \log\left(\frac{255}{RMSE}\right) \text{ db}$$

Where, 255 is the maximum pixel value for an 8 bits/pixel gray-scale image.

The car and lena image is taken for the experimental purpose as shown in Fig.-1 (a) and (b).



Fig. 1(a)

The impulse noise ratio 40%, 70% and 90% are mixed with the car and lena image for experiment. Fig. 2 & 3 shows the comparative qualitative analysis of this proposed algorithm. Here it is compared with the progressive switching median filter (PSM) here median value is replaced whereas after substitution of mid value of sum of mean and median of selected pixels and results shown that it performs better for high noise ratio. Figure 1(b)

5. CONCLUSION

This modified propose filter has a better performance even when the noise ratio is very high. When noise ratio is very less the difference between the PSM and modified version is

approximately same and can not be differentiated on visual basis due to redundancy of good pixels image corrupted with low level of impulse noise. However at the presence of high level of noise it works very well.

REFERENCES

- [1] Digital image processing, R. Gonzalez and R. Woods, PHI II Edition 2008.
- [2] S-J. Ko and Y. H. Lee, "Centre-weighted median filters and their applications to image enhancement" IEEE Trans. Circuits and Syst., vol. 38, pp. 984-993, Sept 1991.
- [3] T. Sun and Y. Neuvo, "Detail-preserving median based filters in image processing," Pattern Recognit. Lett., vol. 15, pp. 341-347, Apr. 1994.
- [4] Zhou Wang and David Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images," IEEE Trans. Circuits & Systems II: Analog & Digital Signal Processing, vol. 46, no. 1, pp. 78-80, Jan. 1999.
- [5] Brownrigg, D.: „The weighted median filter", Comm. Assoc. Comput., 1984, 27, (8), pp. 807–818.
- [6] Yin, L., Yang, R., Gabbouj, M., Neuvo, Y.: „Weighted median filters: a tutorial", IEEE Trans. Circuits Syst., 1996, 43, (3), pp. 157–192
- [7] Zhou, H., Zeng, B., Neuvo, Y.: „Weighted FIR median hybrid filters for image processing". Proc. Int. Conf. on Circuits and Systems, Shenzhen, China, 1991, pp. 793–796
- [8] Ko, J., Lee, J.-H.: „Center weighted median filters and their application to image enhancement", IEEE Trans. Circuits Syst., 1991, 38, (9), pp. 984–993.
- [9] Chen, T., Wu, H.R.: „Adaptive impulse detection using center-weighted median filters", IEEE Signal Process. Lett., 2001, 8, (1), pp. 1–3
- [10] Shuqun Zhang and Mohammad A. Karim "A new impulse detector for switching median filters", IEEE Signal Processing Letters, Vol. 9, No. 11, November 2002.
- [11] A. Fabijanska and D. sakowski, "Noise adaptive switching median based filter for impulse noise removal from extremely corrupted images" IET journal, May-2010

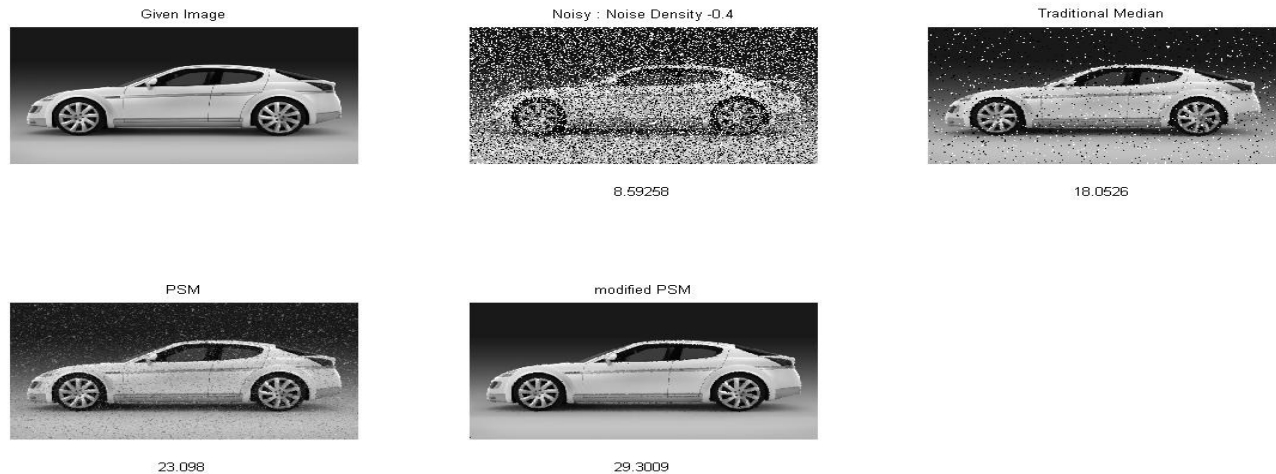


Fig. 2(a)

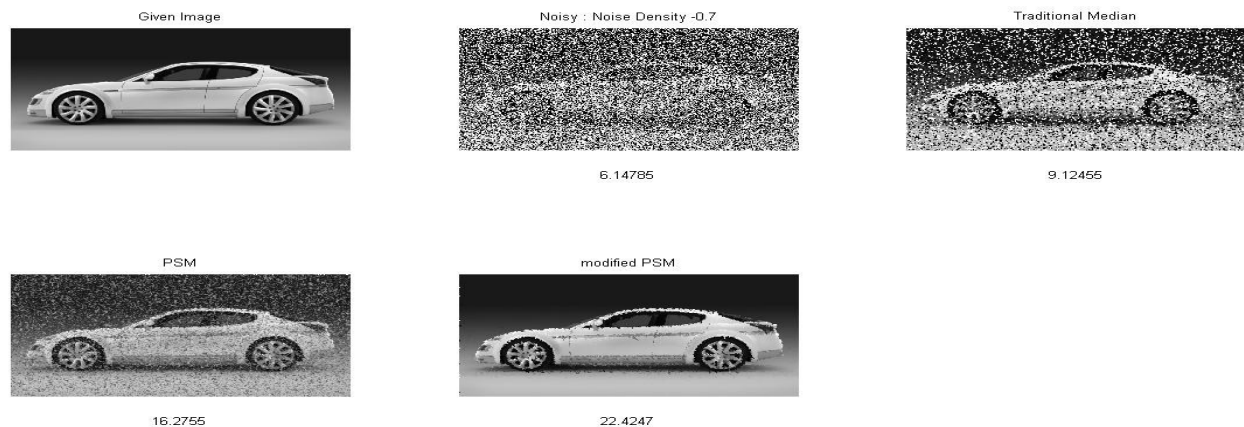


Fig. 2(b)

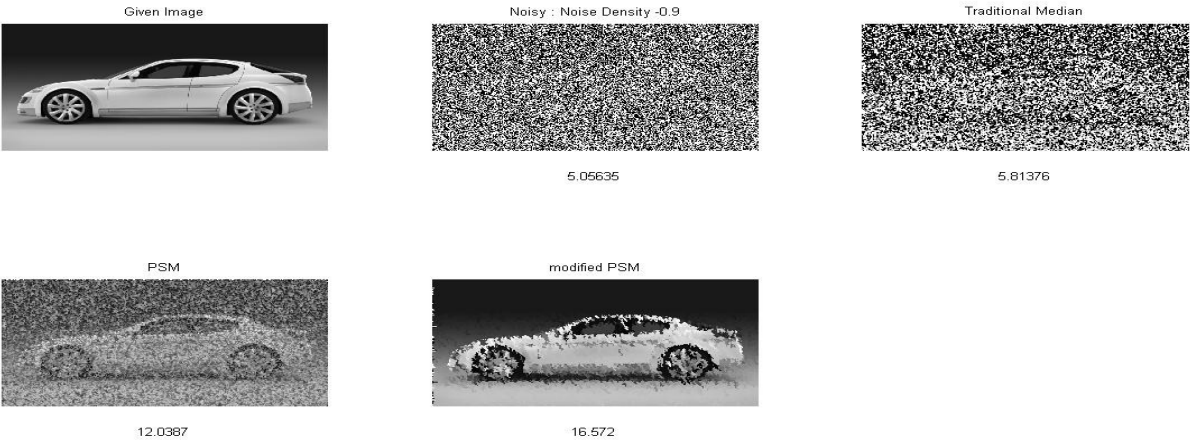


Fig. 2(c)

Fig. 2(a) , 2(b) & 2(c) shows Performance of various Median Filters over a car image mixed with a noise of 40% , 70% and 90% respectively with PSNR value at the bottom of each image



Fig 3(a)



Fig 3(b)

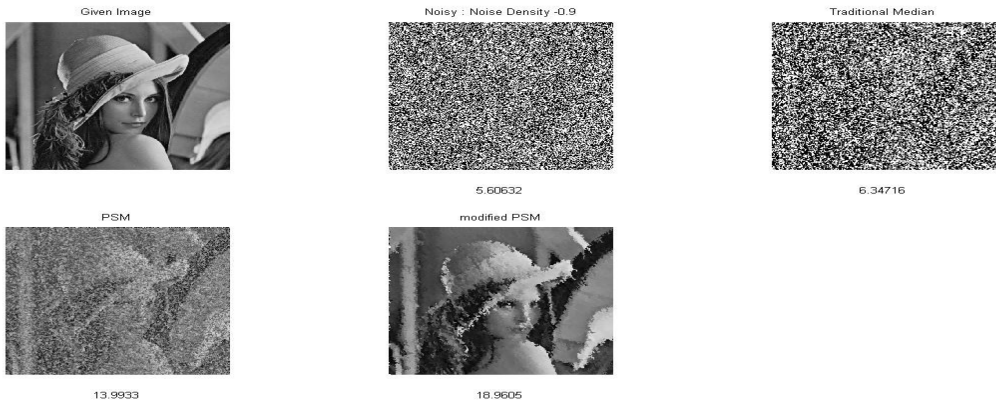


Fig 3(a)

Fig.3(a) , 3(b) & 3(c) shows Performance of various Median Filters over a lena image mixed with a noise of 40% , 70% and 90% respectively with PSNR value at the bottom of each image.

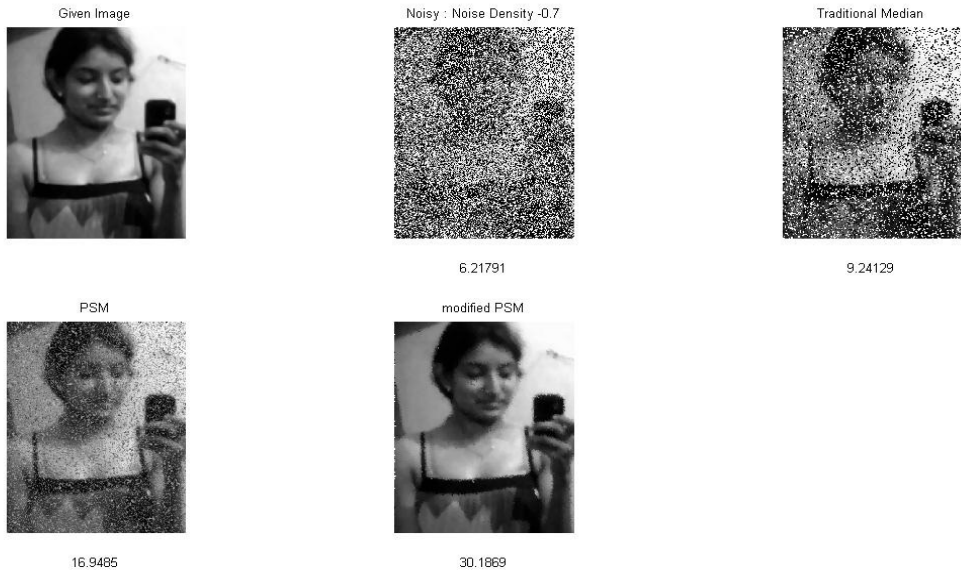


Fig. 4

Fig. 4. Shows Performance of various Median Filters over a Girl image mixed with a noise of 70% with PSNR value at the bottom of each image.

A Survey on Capacity Analysis for Multiuser MIMO Downlink System in Wireless Communication

Abhishek Gupta¹, Garima Saini²

¹ME Student, Department of ECE, NITTTR, Chandigarh

¹abhishek.gupta01@yahoo.com

NITTTR, Chandigarh, Assistant Professor, Department of ECE,

²garimasaini_18@rediffmail.com

Abstract: Very few technologies have shown as much impact on the trajectory of evolution of wireless communication systems as multiple input multiple output (MIMO) systems. MIMO systems have already been applied in the existing 802.11n and 802.16e standards resulting in a huge increase in their achievable rates. A relatively recent idea of extending the benefits of MIMO systems to multi-user scenarios seems promising in the context of achieving high data rates keeping in the mind for future cellular standards after 3G. For applications such as wireless LANs and cellular telephony, MIMO systems will likely to be deployed in environments where a single base station must communicate with many users simultaneously. As a result, the study of multi-user MIMO systems has emerged recently as an important topic for research in future[1]. Such systems have the potential to combine the high capacity that can be achieved with MIMO processing by using space-division multiple access techniques. This paper aims at giving an insight into Multiuser MIMO downlink systems—its concept, capacity, and transmission techniques related issues. In this paper several approaches including linear and non-linear channel precoding are reviewed which analyze the capacity for multiuser MIMO downlink channel. We conclude by describing the future areas of research in multi-user MIMO communications system.

Keywords: Multiuser, MIMO, Sum capacity, Downlink, Precoding

1. INTRODUCTION

Multiple-input multiple-output communication systems have attracted attention in recent years. Such multiple antennas potentially allow higher throughput, increased diversity, and reduced interference as they communicate with multiple wireless users[2]. The use of MIMO systems was initially intended for point to point communication. MIMO techniques were first examined in single-user scenarios. The natural extension of this thought would be to consider MIMO systems in a multi-user scenario. It is well known that in a MIMO system with N_B transmit and N_M receive antennas, capacity grows linearly with $\min(N_B, N_M)$. Current interest in the multiple user case is motivated by recent results indicating that similar capacity scaling applies when an N_B antennas communicates with N_M users. The vision for next generation cellular networks includes data

rates approaching 100 Mb/s for highly mobile users and up to 1 GB/s for low mobile or stationary users. These require efficient use of the existing spectrum. Multiuser MIMO technology is expected to play a key role in this context. There are two challenges in a multi-user MIMO scenario: uplink (where multiple users transmit simultaneously to single base station) and downlink (where the base station transmits to multiple independent users). In order to separate the signals transmitted by the users the uplink challenge is addressed using array processing and multi-user detection techniques by the base station. The downlink challenge is somewhat different [2, 3]. MU-MIMO downlink channel is similar to that of single-user MIMO (SU-MIMO) except that the receiver antennas are distributed among different independent users. In the multi-user case, interference must be taken into account and balanced against the need for high throughput. A transmission scheme that maximizes the capacity for one user in the network might result in unacceptably high interference for the other users, rendering their links useless. If high throughput is the goal, a better approach might be to maximize the sum capacity of the network, or the maximum sum transmission rate, where the inter-user interference is taken into consideration[4]. This paper is organized as follows: An overview of the Multiuser MIMO is presented in Section II. In Section III, capacity analysis of Multiuser MIMO downlink channel is detailed. In section IV various transmission methods for multiuser MIMO downlink channel are discussed. Finally conclusions are drawn with some future work in V.

2. MULTIUSER MIMO

Most of the communication systems deal with multiple users who are sharing the same radio resources. Fig.1 illustrates a typical multi-user communication environment in which the multiple mobile stations are served by a single base station in the cellular system. In Fig.1 users are selected and allocated communication resource such as time, frequency, and spatial stream. Suppose that the base station and each mobile station are equipped with N_B and N_M antennas, respectively. As K independent users form a virtual set of KN_M antennas which communicate with a single BS with N_B

antennas, the end-to-end configuration can be considered as a $(K N_M) N_B$ MIMO system for downlink, or $N_B (K N_M)$ MIMO system for uplink. In multiuser communication system, multiple antennas allow the independent users to transmit their own data stream in the uplink at the same time or the base station to transmit the multiple user data streams to be decoded by each user in the downlink. This results to the increase in degrees of freedom with multiple antennas. In the multi-user MIMO system, broadcast channel (BC) and multiple access channel (MAC), are referred to as downlink and uplink channels respectively. Since all data streams of K independent users are available for a single receiver of the base station in the multiple access channels, the multi-user MIMO system is equivalent to a single user $(K N_M) N_B$ MIMO system in the uplink. Similar to the single-user MIMO system, therefore, it can be shown that the uplink capacity of multi-user MIMO system is proportional to $\min(N_B \times K N_M)$. Therefore efficient use of the existing spectrum is required. MU-MIMO technology is expected to play a important role in this context. This creates a challenge in decoding the received symbols since joint decoding requires each user to have the symbol received from all the receiver antennas of all the users[5]. It is practically not possible to achieve this level of coordination among all users. Almost all of the proposed techniques presented for addressing the MU-MIMO downlink problem employ processing of data symbols at the transmitter itself known as precoding. Although precoding is not a new concept and has been used in SU-MIMO systems as well, it was optional and used only to improve signal to noise ratio (SNR) at the receiver. However, in MU-MIMO systems precoding is essential to remove or minimize inter-user interference. Precoding is performed with the help of downlink channel state information or channel state information (CSI). This requires the transmitter to know the downlink CSI of each user in order to model the precoding transformation variables for each user. A number of different techniques to address the issue of MIMO downlink transmission have been discussed in this paper.

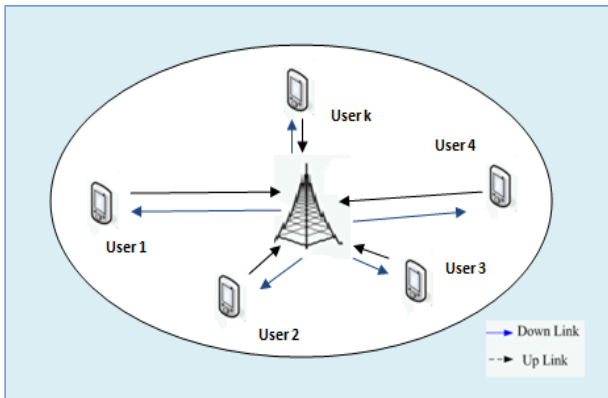


Fig. 1. Multi-user MIMO communication systems

3. CAPACITY IN MULTIUSER MIMO DOWNLINK CHANNELS

Capacity is an important tool for characterizing any communication channel. In a single-user channel, capacity is the maximum amount of information that can be transmitted as a function of available bandwidth in which transmitted power is a constrained. For the multi-user MIMO downlink channel, the problem is somewhat more complex. Given a constraint on the total transmitted power, it is possible to allocate varying fractions of that power to different users in the network, so that a single power constraint can yield many different information rates. The result is a “capacity region” which is illustrated in Fig.2 for a two-user channel. The maximum capacity for user 1 is achieved when 100% of the power is allocated to it.

$$R_k = \frac{\log \left(I + H_k \left(\sum_{j=1}^k S_j \right) H_k^* \right)}{\log \left(I + H_k \left(\sum_{j=1}^k S_j \right) H_k^* \right)}$$

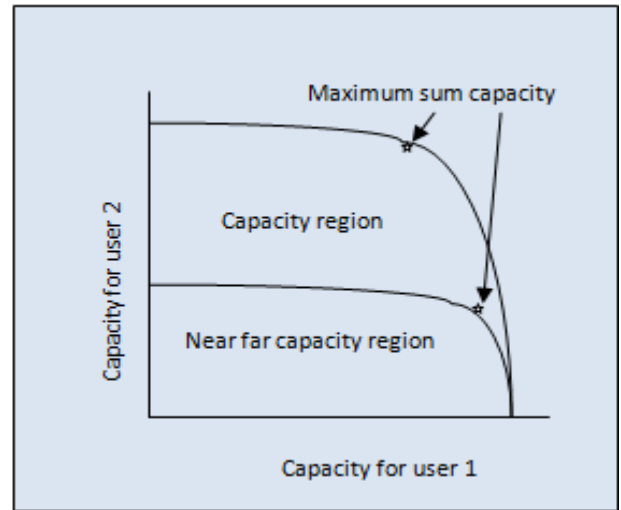


Fig. 2. An illustration of a Multi-user capacity region

When user 2 has all the power, maximum capacity is also obtained in that case [6]. For every possible power distribution in between, there is an achievable information rate, which results in the capacity regions shown in the illustrated region. Two regions are shown in Fig.2 the bigger one for the case where both users have roughly the same maximum capacity and the other for a case where they are different (due to the fact that user 2's channel being attenuated compared to user 1). For K users, the capacity region is characterized by a K -dimensional region. The maximum achievable throughput of the system is characterized by the point on the curve that maximizes the sum information rates of all of the users and is referred to as the sum capacity of the channel. This point is shown in Fig.2 by asterisks. One example where this may be the case is

when the “near-far” problem occurs, where one user has a strongly attenuated channel compared to other users. As shown in Fig.2, obtaining the sum capacity would come at the cost of the user with the attenuated channel. The sum capacity for a system described by Fig.2[6] has been formulated using the dirty paper coding (DPC) framework (see, for example for the case of Gaussian noise. The capacity is defined in terms of the achievable rate for each user given the set of covariance matrices for each transmitted data vector $S_k = \xi\{s_k s_k^*\}$:

4. TRANSMISSION TECHNIQUES IN MULTIUSER MIMO DOWNLINK CHANNELS

The main difficulty in data transmission in Broadcast channel is that the coordinated signal detection on the receiver side is not straightforward, and thus, interference cancellation at BS is required. There are various linear precoding and non linear precoding techniques for multiuser MIMO downlink transmission which solves this problem.

A. Linear Precoding Techniques

A simple way of dealing with inter user interference is by imposing the constraint that all interference terms be zero. Assuming that $N_M \leq N_B$, this can be accomplished at the transmitter by precoding D which is the transmit data vector with the pseudo inverse of the channel matrix: $x = H^\dagger D = H^* (HH^*)^{-1} D$. At the receivers, this approach results in $y = D + w$. This technique is referred to as channel inversion, for the case where H is square. The columns of H^\dagger can be weighted to give different signal to noise ratio for each user, depending on their given rate requirement. *Channel inversion* is a good technique for low-noise or high-power situations. But it does not result in the linear capacity growth with $\min(N_M, N_B)$ that should be achievable in the multi-user channel. This is because with a power constraint, an ill-conditioned channel matrix when inverted will require a large normalization factor that will dramatically reduce the SNR at the receivers[7]. The drawbacks of channel inversion are due to the strict requirement that the interference at the receivers should be identically zero. A limited amount of interference at each receiver allows them to consider a larger set of potential solutions that can potentially provide higher capacity for a given transmit power level, or a less transmit power for a given rate point. This behaviour is seen in the solutions that maximize sum capacity; they allow some level of MAI (multiple access interference) at each receiver. One simple approach for this derives from linear minimum mean squared error (MMSE) receivers used in the uplink. If we assume white noise and power constraint P , the MMSE uplink receiver is given by $(H^* H + K/P)^{-1} H^*$, where H is the uplink channel. For the down-link, it is possible to assume a similar MMSE-like structure, using $x = H^* (H^* H + \alpha I)^{-1} D$. This type of “regularized” channel inversion shows that the loading factor $\alpha = K/P$ maximizes the signal-to-

interference-plus-noise ratio (SINR) at the receiver when this scheme is used. This simple procedure results in a solution that does achieve linear growth in throughput with $\min(N_M, N_B)$, but at a rate that is somewhat slower than that for capacity. Both types of channel inversion described here are designed to achieve some signal to noise ratio that is identical for each user. Therefore in next generation communication systems there will be an increasing need to support heterogeneous wireless services, which implies that each user may have different bandwidth and/or signal to noise ratio requirements. One way to achieve this requirement is to adjust the amount of transmitted power to each user. This is straightforward with direct channel inversion because the sub channels created to each user are independent, but with regularized inversion, changing the power transmitted to one user changes the interference for all other users. This necessitates a *beam forming* solution where the beam forming vectors and power weights are jointly optimized user may have different bandwidth and/or SINR requirements [8]. This problem can be extended to consider cases where the users also have arrays, a scenario of interest for next-generation systems. Adding multiple antennas at each receiver makes it possible for the transmission of parallel data streams to multiple users. Channel inversion in this case is not an efficient solution, as forcing two closely spaced antennas belonging to a single user to receive different signals would require extra power when the channels for these antennas are highly interrelated. One approach to this problem is to use block channel inversion or *block diagonalization*. This approach is essentially a generalization of channel inversion that optimizes the power transfer to a group of antennas rather than a single antenna[9]. Same as channel inversion for single antennas, this approach requires that the number of transmit antennas should be larger than the total number of receive antenna and does it not only achieves capacity, but also offers relatively less computational cost.

B. Non Linear Precoding Techniques

We now turn to a nonlinear technique based on the concept of “writing on dirty paper” introduced by Costa[10] in which the traditional additive Gaussian noise channel is modified to include an additive interference term that is known at the transmitter: received signal = transmitted signal + interference + noise. The simplest thing to do in such an environment would be to set the transmitted signal equal to the desired data minus the interference, but such an approach requires increased power. Costa proved that the capacity of this channel is the same as if the interference was not present and therefore no more power is required to cancel the interference which is used in an additive Gaussian noise channel. To use Costa’s analogy, writing on dirty paper is information i.e. theoretically equivalent to writing on clean paper when one knows in advance where the dirt is. Costa’s approach is theoretical and does not provide a

practical technique for meeting capacity solutions. As the transmitter has channel state information, it knows what interference user 1's signal will produce at user 2, and hence can design a signal for user 2 that avoids the known interference. This concept has been used to characterize the sum-capacity and capacity region of the multi-antenna multi-user channel. The most well-known dirty paper technique for the MIMO downlink uses a QR decomposition of the channel, which can be written as the product of a lower triangular matrix L with a unitary matrix $H = LQ$. The signal to be transmitted is precoded with the Hermitian transpose of unitary matrix Q , resulting in the effective channel L . In this case no interference is seen from other users and signal can be chosen without regard for the other users. The second user sees interference only from the first user and can be overcome by using dirty paper coding. Subsequent users are dealt with in a similar case. Another approach applies dirty paper techniques directly, rather than for individual users[11]. An important difference between the multi-user MIMO channel and the interference channels for which non linear techniques like dirty paper are designed is that the interference depends on the signal being designed. In the previous section this problem is solved using QR-type decomposition, so the interference for any particular user depends only on the interference generated by previous users. Dirty paper coding is therefore applied to cancel this interference. An alternate technique is to design all the signals jointly and this approach took matrix algebra to solve for the signal to be transmitted.

The simple dirt paper technique of applying a modulo operation to the transmitted and received data is shown to operate close to the sum capacity vector precoding. It is the modification of channel inversion, where the desired signal D is offset by a vector l of integer values chosen to minimize the power in the transmitted signal, $x = H^{-1} (D + \tau l)$: where τ is chosen in the same way as for the successive algorithm described above. As with basic channel inversion, this encoding results in the k^{th} receiver seeing an additive Gaussian channel $y_k = D_k + \tau l_k + w_k$. The integer offset l_k is removed by applying a modulo function at the receiver, resulting in a signal that appears to be very much like an additive noise channel. The modification of this technique uses a regularized inverse at the encoder rather than simple channel inversion. The transmitted signal in this case is $x = H^*(HH^* + K/\rho)^{-1} (D + \tau l)$, where the vector l is again chosen to minimize the norm of x . Decoding occurs in the same way as for the non regularized approach. Another approach to this includes use of the lattice techniques for dirty paper coding. The lattice used is the simple one dimensional lattice defined by the modulo function.[12]. Fast algorithms for finding the integer vector l have been proposed is based on lattice reduction and the VBLAST algorithm. These techniques have lower complexity than the sphere-algorithm based techniques. Only single-antenna users are considered as a simple extension to situations

involving multiple receive antennas per user to treat each antenna as a different user. At low SNR, precoding with the regularized inverse channel performs the best results, while at high SNR the regularized precoding technique is best for the same performance. The basic and regularized vector precoding techniques have a significant diversity advantage over the other techniques in this uncoded example. One possible explanation for basic regularized inversion's performing better than regularized vector precoding at low SNR is that the cubical lattice used in the latter algorithm is finite. Lattices as described in may enable the vector precoding techniques to perform as well as the basic inversion-based methods[13]. DPC on the transmitter side is very similar to decision feedback equalization (DFE) on the receiver side. In fact, combination of DPC with symmetric modulo operation turns out to be equivalent to *Tomlinson Harashima (TH)* precoding. TH precoding was originally invented for reducing the peak or average power in the decision feedback equalizer (DFE), which suffers from error propagation[14]. The original idea of TH precoding in DFE is to cancel the post-cursor ISI in the transmitter, where the past transmit symbols are known without possibility of errors. In fact, it requires a complete knowledge of the channel impulse response, which is only available by a feedback from the receiver for time-invariant or slowly time-varying channel.

5. CONCLUSION

The multi-user MIMO problem has recently started to attract the attention of the researchers. In this paper we have presented a brief overview of two classes of downlink transmission algorithms: linear processing techniques and non linear processing techniques. Linear techniques are simple as compared to non linear techniques and relatively cheap computationally, but they are not able to reach the sum-capacity of the channel. Techniques based on dirty paper coding perform much better and approach the theoretical limits of the channel, but require complicated coding schemes. Combination of both these techniques using various different methods can be explored in future. Higher dimensional lattices could be used to further approach the sum-capacity of the multi-user channel. The most unsolved problem in this field which is clearly visible has been determining the capacity region for the MIMO multi-user channel. Though a solution was recently found to the Gaussian problem, there are many problems yet to be solved (e.g. in the non-Gaussian case). An analysis of the penalty for using imperfect or outdated feedback of channel information would be of significant benefit to system designers. The sum-capacity when only the transmitter or when no one knows the channel would also provide insight for practical coding schemes. A related area of research is analysis of a system where the transmitter and/or receiver know only the statistics of the channel coefficients.

REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, pp. 425-495, 2005.
- [2] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An Introduction to the Multi-user MIMO downlink link," *IEEE Communication Magazine*, vol. 42, no. 10, pp. 60-67, 2004.
- [3] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936-3964, Sep. 2006.
- [4] Yong Soo Cho, Jaekwon Kim, Won Young Yang, Chung Gu Kang, "MIMO-OFDM Wireless Communication with Matlab," *IEEE Press*, pp. 209-236, 2010.
- [5] Jinkyu Kang, Keonkook Lee, Jungho Myung, Joonhyuk Kang, "Multiuser MIMO Downlink with Linear Precoding for Full Multiplexing Gain", *IEEE Conference on Vehicular Technology Fall*, vol. 72, pp. 1-5, Sept. 2010.
- [6] Christian B. Peel, Quentin H. Spencer, A. Lee Swindlehurst, Martin Haardt and Bertrand M. Hochwald, "Linear and Dirty Paper techniques For the Multi-User MIMO Downlink", *Space-Time Processing for MIMO Communication*, pp. 209-236, 2005.
- [7] Youxiang Wang, Soojung Hur, Yongwan Park, and Jeong-Hee Choi, "Efficient User Selection Algorithms for Multiuser MIMO Systems with Zero-Forcing Dirty Paper Coding", *Journal of Communication and Networks*, vol. 13, no. 3, pp. 232-239, June 2011.
- [8] Chan-Byoung Chae, Seijoon Shim, Robert W. Heath, "Block Diagonalized Vector Perturbation for Multiuser MIMO Systems", *IEEE Trans. on wireless Communication*, vol. 7, no. 11, pp. 4051-4057, Nov. 2008.
- [9] Wibowo Hardjawana, Branka Vucetic, Yonghui Li, "Multi-User Cooperative Base Station Systems with Joint Precoding and Beamforming", *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 6, pp. 1079-1091, Dec. 2009.
- [10] M. H. M. Costa, "Writing on dirty paper", *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.
- [11] Liang Sun, and Matthew R. McKay, "Eigen-Based Transceivers for the MIMO Broadcast Channel with Semi-Orthogonal User Selection", *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5246-5261, Oct. 2010.
- [12] Christoph Windpassinger, Robert F. H. Fischer, Tomá s Vencel, Johannes B. Huber, "Precoding in Multiantenna and Multiuser Communications", *IEEE Trans. on Wireless Communication*, vol. 3, no. 4, pp. 1305-1315, July 2004.
- [13] D. Gesbert, M. Kountouris, R.W. Heath Jr., Chan Byoung Chae, "Shifting the MIMO Paradigm: from single user to multiuser communications," *IEEE Signal Processing Mag.*, pp. 36-46, Sep. 2007.
- [14] Fernando Pérez-Cruz, Miguel R. D. Rodrigues and Sergio Verdú, "MIMO Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1070-1084, March 2010.

Channel Performance by using Adaptive Equalization Techniques in MIMO System for Multipath Fading Environment

Geetesh Kwatra¹, Liladhar Malviya²

¹Dept of Elect. and Telecomm, CDSE, Indore, India, kumar.geetesh@gmail.com

²Dept of Elect. and Telecomm, SGSITS, Indore, India, ldmalviya@rediffmail.com

Abstract: Wireless communication user endlessly demands for higher spectral rate and better reliability of the system. To realize these demands, wireless communication engineer faced challenging problems i.e. signal fading, delay spread etc. It can greatly impair the performance of a data communication system channel's inherent complexity such as time varying nature of the channel. In this paper we address increased channel performance (by adaptive equalizer (maximum likelihood estimator in Gaussian channel) and channel capacity via multiple input multiple outputs (MIMO)). These MIMO systems are considered in time varying channel and information available in form of channel fade level is tracked by both transmitter and receiver. We also focus on get maximize channel capacity in MIMO system with different CSI configuration by the singular value decomposition (SVD).

Keyword: MIMO, SVD, ML, CSI.

1. INTRODUCTION

We used single-user communication model and consider a point-to-point link where the transmitter is equipped with n_T antennas and the receiver employs n_R antennas. Single user assumption in the depiction as point-to-point link, we suppose that no inter-symbol interference (ISI) occurs. This implies that the bandwidth of the transmitted signal is very small and can be assumed frequency-flat (coherent bandwidth), so that each signal path can be represented by a complex-valued gain factor. For practical purposes, it is common to model the channel as frequency-flat whenever the bandwidth of the system is smaller than the inverse of the delay spread of the channel; hence a wideband system operating where the delay spread is fairly small.

Now let $h_{i,j}$ be the complex-valued path gain from transmit antenna j to receive antenna i (the fading coefficient). If at a certain time instant the complex-valued signals $\{s_1, \dots, s_{n_T}\}$ are transmitted via the n_T antennas, respectively, the received signal at antenna i can be expressed as

$$y_i = \sum_{j=1}^{n_T} h_{i,j} s_j + n_i \quad (1)$$

where n_i represents additive noise this linear relation can be easily written in a matrix framework. Thus, let \mathbf{s} be a vector of size n_T containing the transmitted values, and \mathbf{y} be a vector of size n_R containing the received values, respectively. We have $\mathbf{s} \in \mathbb{C}^{n_T}$, $\mathbf{y} \in \mathbb{C}^{n_R}$. [3]

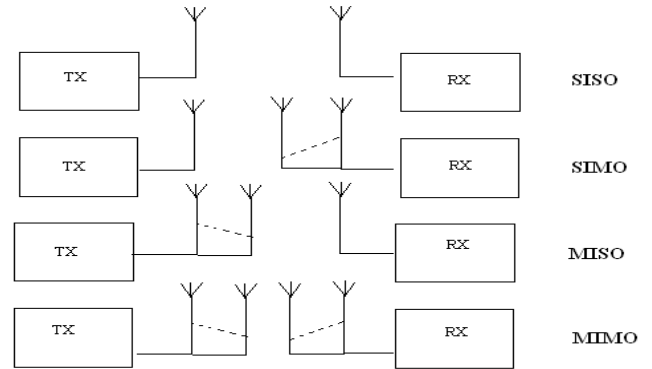


Fig. 1. MIMO channel model

2. CHANNEL CAPACITY AND SINGULAR VALUE DECOMPOSITION

A definition of the channel capacity is made and there is also a discussion about how the capacity varies with the number of transmitting and receiving antennas. An interesting discussion on how many transmitting and receiving antennas one should use can be found in. Intended for this there are tradeoff between number of antenna (say channel capacity) and model complexity. For hands on this computation we used singular value decomposition (SVD), the channel matrix \mathbf{H} can be decomposed into a product of three matrices as follows.

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \quad (2)$$

Where \mathbf{U} are a unitary matrix of dimension $n_R \times n_R$, \mathbf{V} are also a unitary matrix of dimension $n_T \times n_T$ and $\mathbf{\Sigma}$ is a $n_R \times n_T$ matrix whose elements are all zero except for the diagonal where there will be $\min(n_T, n_R)$ of the \mathbf{H} matrix eigenvalues.

The V^H represents the Hermitian matrix. The three matrices correspond to three different steps:

1. Projection into Tx Eigenmodes: Each of the first $\min(n_T, n_R)$ columns in V matrix is a unit-norm vector corresponding to each of the transmit Eigenmodes. The relative phases and amplitudes between transmit elements required to excite each Eigenmode are described by each of these column vectors. The first $\min(n_T, n_R)$ components of the vector $V1$ are therefore the projection of the transmit vector x into the transmit eigenvector subspace. When $n_T > n_R$, the remaining $n_T - n_R$ components belong to the subspace orthogonal to the transmit eigenvector subspace and cannot influence the received vector.

2. Weighting by singular values:

Each of the $\min(n_T, n_R)$ components corresponding to the transmit eigenmodes are weighted by its associated singular value contained in the main diagonal of the matrix Σ .

3. Mapping into Rx eigenmodes:

Each of the first $\min(n_T, n_R)$ columns in U matrix is a unit-norm vector corresponding to the mapping of each eigenmode in the receiving space. The remaining $N - \min(n_T, n_R)$ vectors are not Rx eigenmodes because they cannot be excited by the transmitter. The discussion above leads to some interesting conclusions:

- If $n_T > n_R$, some power is wasted on exciting a subspace orthogonal to the receiver. The receiver cannot interpret this subspace and it's totally unnecessary. If the power is allocated uniformly over the transmitter there will be an average power loss of $10 \log_{10} \frac{n_T}{n_R}$.
- If $n_T < n_R$, there is no power loss. There will be $n_T - n_R$ dimensions in the receiver space, which are not excited by the receiver eigenvectors.

3. MAXIMUM CAPACITY, IF THERE IS NO CSI AT THE TRANSMITTER

The capacity for a MIMO channel will be defined. There will also be shown that by supplying the transmitter with channel state information (CSI) it is possible to greatly increase the capacity. The channel is estimated in the receiver through to use of a short known transmitted training sequence. For a one-antenna system one can apply the Shannon formula:

$$C = B \log_2 \left(1 + \frac{S}{N} \right) \quad (3)$$

Where C is the channel capacity and is measured in [Bits/(sec*Hz)], B is the bandwidth and S/N is the signal-to-

noise ratio. This is the maximum rate the channel can give with arbitrary low probability of bit errors (allowing infinite coding delay). Hence, it is an upper limit on the practical achievable bit-rate. When one uses a MIMO system one have to use a generalized version of Shannon's formula:[13]

$$C = \log_2 \left(\det \left(I + \frac{\rho_k H H^*}{\sigma^2} \right) \right)$$

$$C = \sum_{k=1}^n \log_2 \left(I + \frac{\rho_k \lambda_k}{\sigma^2} \right) \quad (4)$$

where H is the transition matrix and $(*)$ denotes the complex conjugate transpose. ρ_k , $k=1,2,\dots,n$ is the transmitted energy through channel "mode" k with power gain λ_k , as a consequence of this

$$\sum \rho_k = P_T \quad (5)$$

here P_T is the totally available power. λ_n is also the eigenvalue to $H H^*$ where $\lambda_1 > \lambda_2 > \dots > \lambda_n$. And n is $\min[n_T = \text{number of transmitters}, n_R = \text{number of receivers}]$.

4. MAXIMUM CAPACITY, IF THE CHANNEL IS KNOWN TO THE TRANSMITTER

A method to use CSI at the transmitter to transmit in an optimal way will be explained. It is called the Water filling technique and it is proven for a 2x2 system. The receiver can gain knowledge about the channel by the use of a known training sequence but if the transmitter should know anything about the channel it is necessary to use a feedback channel. The feedback channel consumes bandwidth in the channel or alternatively the capacity will decrease but this will be ignored in this discussion.

When the transmitter knows the Eigenvalues and eigenvectors corresponding to the H matrix and the noise power (σ^2) it can use this information to transmit in a smarter way. See for an excellent description of the "Water filling technique". The Water filling technique is used to determine the powers ρ_k transmitted in each channel to achieve to greatest possible capacity. Consider a MIMO communication link with a shared total power budget of P_T , the capacity is then accordingly equation to (3)

$$C = \sum_{k=1}^n \log_2 \left(I + \frac{\rho_k \lambda_k}{\sigma^2} \right) \quad (6)$$

To achieve the greatest possible capacity ρ_k should be chosen in such a way that for every mode k

$$\rho_k = \left(\mu - \frac{\sigma^2}{\lambda_k} \right)^+ \quad (7)$$

where μ is the “water level”. Furthermore μ should be chosen such that the total power budget is not exceeded, that is equation (4) $\sum \rho_k = P_T$.

Important note: To obtain the optimal capacity the transmitter must have perfect knowledge of the H matrix (the eigenvalues and eigenvectors of H) and σ^2 . There will not be a general proof of the Water filling technique, but the idea will be shown for a 2x2 MIMO system. The Method of Lagrange Multipliers will be used, Maximize $f(\rho_1, \rho_2)$ [14]

$$\begin{aligned} \text{Where } f(\rho_1, \rho_2) &= \log_2 \left(1 + \frac{\rho_1 \lambda_1}{\sigma^2} \right) + \log_2 \left(1 + \frac{\rho_2 \lambda_2}{\sigma^2} \right) \\ &= \log_2 \left(1 + \frac{\rho_1 \lambda_1}{\sigma^2} + \frac{\rho_2 \lambda_2}{\sigma^2} + \frac{\rho_1 \rho_2}{\sigma^2 \sigma^2} \lambda_1 \lambda_2 \right) \end{aligned}$$

under the power constraint $(\rho_1, \rho_2) = \rho_1 + \rho_2 - P_T = 0$

which is an equivalent problem with Maximize $f(\rho_1, \rho_2) = \left(1 + \frac{\rho_1 \lambda_1}{\sigma^2} + \frac{\rho_2 \lambda_2}{\sigma^2} + \frac{\rho_1 \rho_2}{\sigma^2 \sigma^2} \lambda_1 \lambda_2 \right)$

under the constraint $(\rho_1, \rho_2) = \rho_1 + \rho_2 - P_T = 0$

since the $\log_2(\dots)$ function is monotonic. [13]

Let $L(\rho_1, \rho_2, v) =$

$$\left(1 + \frac{\rho_1 \lambda_1}{\sigma^2} + \frac{\rho_2 \lambda_2}{\sigma^2} + \frac{\rho_1 \rho_2}{\sigma^2 \sigma^2} \lambda_1 \lambda_2 \right) + v(\rho_1 + \rho_2 - P_T).$$

For critical points we want

$$0 = \frac{\partial L}{\partial \rho_1} = \frac{\lambda_1}{\sigma^2} + \frac{\rho_2}{\sigma^2 \sigma^2} \lambda_1 \lambda_2 + v \quad (8)$$

$$0 = \frac{\partial L}{\partial \rho_2} = \frac{\lambda_2}{\sigma^2} + \frac{\rho_1}{\sigma^2 \sigma^2} \lambda_1 \lambda_2 + v \quad (9)$$

$$0 = \frac{\partial L}{\partial v} = \rho_1 + \rho_2 - P_T \quad (10)$$

$$\text{Equation (7)} \geq \rho_2 = -v \frac{\sigma^2}{\lambda_1 \lambda_2} - \frac{\sigma^2}{\lambda_2}$$

$$\text{Equation (8)} \geq \rho_1 = -v \frac{\sigma^2}{\lambda_1 \lambda_2} - \frac{\sigma^2}{\lambda_1}$$

But since V is a variable which can be chosen arbitrary, a substitution can be made without any loss of generality

$$\mu = -v \frac{\sigma^2}{\lambda_1 \lambda_2}$$

and since $\rho_1, \rho_2 \geq 0$

must we have constraints on the choice of μ . In the end we get that by choosing μ in a proper way that satisfies

$$\rho_T = \left(\mu - \frac{\sigma^2}{\lambda_1} \right)^+ + \left(\mu - \frac{\sigma^2}{\lambda_2} \right)^+$$

and an optimal capacity is achieved. Accordingly to the water-filling technique gives three different kinds power allocation depending on the SNR. [14]

5. MAXIMUM LIKELIHOOD ALGORITHM (ADAPTIVE EQUALIZER)

We first obtain the structure of the optimum detector (using matched filter) for digital data transmission for band limited and AWGN model. Let received signal for equivalent low pass filter is

$$r(t) = \sum_n h(t - nT) + z(t) \quad (11)$$

where $z(t)$ represent AWGN. We concern optimum detector can be realize as filter matched to channel filter (low pass filter). Using karhunen – Loeve Fourier series expansion we rewritten our received signal

$$r_{kl}(t) = \lim_{N \rightarrow \infty} \sum_{k=1}^N r_k \phi_k(t) \quad (12)$$

where $\phi_k(t)$ is complete set of orthogonal function and r_k are the observable random variable obtained by projecting $r_{lk}(t)$ on to $\phi_k(t)$ as

$$r_k = \sum_n I_n h_{kn} + z_k \quad k = 1, 2, 3 \dots \quad (13)$$

Where h_{kn} is the value obtained from projecting $h(t - nT)$ onto $\phi_k(t)$. The sequence z_k have Gaussian with zero mean and covariance is $E(z_k^* z_m) = 2 N_0 \delta_{km}$. Now The joint probability density function of the random variable $r_N = [r_1, \dots, r_N]$ condition on the transmitted sequence

$I_N = [I_1, \dots, I_N]$, so $P \leq N$ is

$$P(r_N / I_P) = \left(\frac{1}{2\pi N_0} \right)^N e^{\left(-\frac{1}{2N_0} \sum_{k=1}^N |r_k - \sum_n I_n h_{kn}|^2 \right)}. \quad (14)$$

In the limit as the number N of observable random variables approaches infinity, the logarithm of $P(r_N / I_P)$ is proportional to the the matrix $PM(I_P)$, define as

$$\begin{aligned} PM(I_P) &= - \int_{-\infty}^{\infty} \left| r_1(t) - \sum_n I_n h(t - nT) \right|^2 dt \\ &= \int_{-\infty}^{\infty} |r_1(t)|^2 + 2 \text{Re} \, dt \sum_n [I_n^* \int_{-\infty}^{\infty} r_1(t) h^*(t - nT)] \end{aligned}$$

$$-\sum_n \sum_m [I_n^* \int_{-\infty}^{\infty} r_1(t) h^*(t-nT) h(t-mT) dt] \quad (15)$$

The maximum likelihood estimates of the symbols I_1, I_2, \dots, I_n are those that maximize this quantity. The other integral of involve $r(t)$ give rise to the variable $y_n = y(nT) = \int_{-\infty}^{\infty} r_l(t) h^*(t-nT) dt$

This variable can be generated by passing $r(t)$ through a filter matched to $h(t)$ and sampling the output at the symbol rate $1/T$. The samples $\{y_n\}$ from a set of sufficient statistics for the computation of PM (I_p) or equivalently, of the correlation metrics

$$CM(I_n) = 2\text{Re}(\sum_n I_n^* y_n) - \sum_n \sum_m I_n^* I_m x_{n-m} \quad (16)$$

The metrics that are computed for the ML algorithm for the sequence $\{I_k\}$ are given by equation (15). It can be seen that matrix can be computed recursively in the Viterbi algorithm, according to relation

$$CM(I_n) = CM_{n-1}(I_{n-1}) + \text{Re}[I_n^* (2y_n - x_0 I_n - 2 \sum_{m=1}^L x_m I_{n-m})] \quad (17)$$

The matrix is computed for ML of the sequence.

6. SIMULATION RESULTS AND PARAMETER

A particular MIMO system was modeled with the cooperation of MATLAB in Order to get exact power allocation results. Performance results and plots are mainly based on analytical calculations (Shannon's capacity), however some system parameters - like, fading channel coefficients, and fairness vector of users were randomly generated. For analyzing performance of capacity we calculate parameter value for formation of analytical table as below:

7. SIMULATION RESULT OF CHANNEL CAPACITY WITH DIFFERENT ANTENNA ARRAY

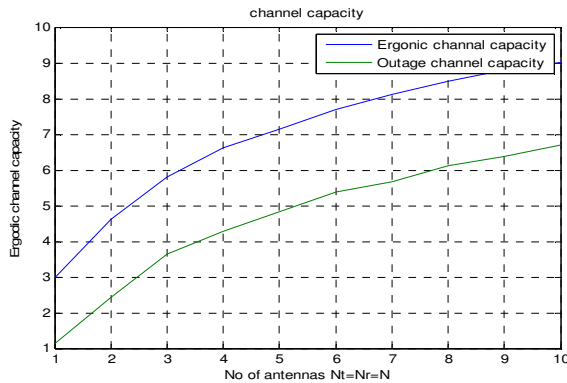


Fig. 2. Performance of Ergodic and outage

Table:1 Ergodic and outage channel capacity with different antenna array

$n_T = n_R = n$	1	2	3	4	5	6
ERGODIC	2.982	4.614	5.805	6.616	7.149	7.686
OUTAGE	1.155	2.441	3.632	4.268	4.839	5.384

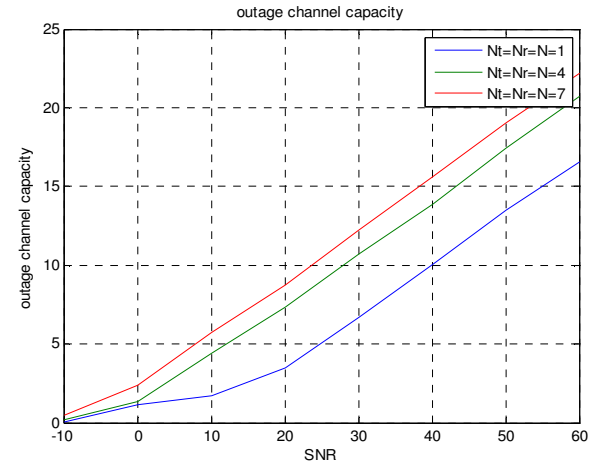


Fig. 3. Performance of Outage channel capacity

Table 2: Outage channel capacity with different antenna array

SNR(in dB)	-10	0	10	20	30	40
$n_T = n_R = n = 1$	0.016	1.139	1.047	3.491	6.691	10.036
$n_T = n_R = n = 4$	0.208	1.35	4.386	7.3773	10.7259	13.8491

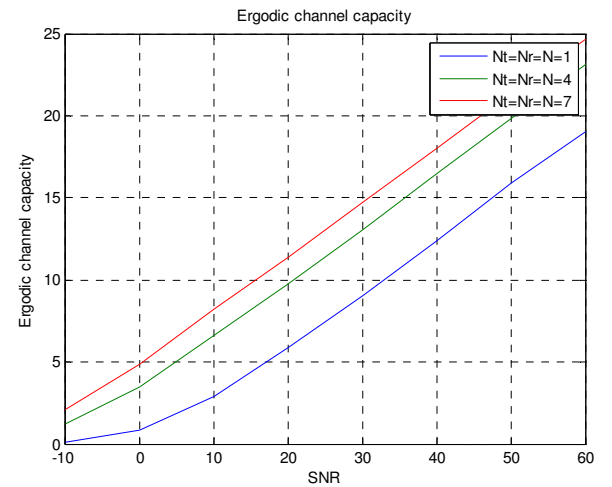


Fig. 4. Performance of Ergodic channel capacity

Table: 3 Ergodic channel capacity with different antenna array

SNR(in dB)	-10	0	10	20	30	40
$n_T = n_R = n = 1$	0.128	0.865	2.889	5.855	12.386	15.866
$n_T = n_R = n = 4$	1.1801	3.4538	6.6351	9.7943	13.0843	16.47

8. CONCLUSION

The capacity of an $n_T \times n_R$ i.i.d. fading MIMO channel H with receiver CSI is at high SNR, the capacity is approximately equal (up to an additive constant) to $n \min \log \text{SNR}$ bits/s/Hz. At low SNR, the capacity is approximately equal to $n \text{SNR} \log 2 e$ bits/s/Hz, so only a receive beam forming gain is realized. With $n_T = n_R = n$, the capacity can be approximated by $n c \text{SNR}$, where $c \text{SNR}$ is the constant. An $n \times n$ MIMO channel, the capacity increases linearly with n over the entire SNR range.

REFERENCES

- [1] Yang Wen Liang "Ergodic and Outage Capacity of Narrowband MIMO Gaussian Channels" 2004.
- [2] MIMO Wireless Communications by Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, H. Vincent Poor Cambridge University Press 2007.
- [3] Martin Wrulich January 11, 2006 "Capacity Analysis of MIMO Systems".
- [4] Gesualdo Scutari, Member, IEEE, Daniel P. Palomar, Member, IEEE, and Sergio Barbarossa, Member, IEEE "The MIMO Iterative Waterfilling Algorithm".
- [5] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity Limits of MIMO Channels. IEEE J. Sel. Areas Commun. 21(5):684-702, June 2003.
- [6] Kenneth W. Shum, Member IEEE, Kin-Kwong Leung, Member IEEE, and Chi Wan Sung, Member IEEE "Convergence of Iterative Waterfilling Algorithm or Gaussian Interference Channels 2007".
- [7] Gesualdo Scutari, member IEEE, Daniel P. Palomar, member, IEEE May 2009 "Convergence of Iterative Water filling Algorithm for Gaussian Interference Channels".
- [8] Md. Noor-A-Rahim1, Md. Saiful Islam "Performance Analysis of MIMO-OFDM System Using Singular Value Decomposition and Water Filling Algorithm" International Journal of Advanced Computer Science and Applications 2011.
- [9] Fundamentals of Wireless Communication by David Tse and Pramod Viswanath Cambridge University press 2005.
- [10] Andrea Goldsmith "Wireless communication" 2005 Cambridge university press.
- [11] Andreas F. Molisch "Wireless communication" 2005 Wiley India publication.
- [12] Yong Soo Cho, Chung G. Kang MIMO OFDM Wireless Communications with MATLAB Wiley publication 2010
- [13] Ezio Biglieri Robert Calderbank Anthony Constantinides Andrea Goldsmith MIMO Wireless Communications Cambridge University Press 2007.
- [14] F. Boixadera Espax, J. Boutros "Capacity considerations for wireless MIMO channels." http://www.com.enst.fr/publications/publi/MMT99_boixadera.ps. 2002-01-08.
- [15] T. Svantesson, A. Ranheim "Mutual coupling effects on the capacity of multielement antenna systems," Acoustics, Speech, and Signal Processing, 2001. Proceedings. 2001 IEEE International Conference on, Volume: 4, 2001.
- [16] John Proakis, Masoud Salehi "Digital communication" 5th edition 2007 McGraw-Hill Companies, Incorporated, 2007.

Medical Image Retrieval Using Texture Features

A. Swarnambiga¹, S.Vasuki², A. Anantha Raja³

¹Ph.D research scholar, ²HOD &Professor, Velammal College of Engineering &Technology, Viraganoor, Madurai-09

³PG Student, College of Engineering guindy, Anna university, Chennai³

¹aswarnambiga@gmail.com, ²hello_vasuki@yahoo.co.in, ³dspananth@gmail.com

Abstract: In this work, we focus on automatic extraction using texture features for medical image retrieval. Features are obtained by computing the energy of image. Two ways of approaches were used to extract the texture features. The first approach of texture feature is obtained by using gray-level co-occurrence matrix (GLCM). The second approach extract features using Haar wavelet. Image retrieval for both the approach achieved by Euclidean distance classifier. The approaches were experimented on two types of modalities i.e. Anatomical modality database A1 with 300 images and Functional modalities database with 300 images. Experimental results indicate that wavelet based retrieval outperforms GLCM based retrieval. Wavelet based method improves retrieval rate from 40 % to 80 % for Anatomical modality and from 30 % to 40 % on Functional modality in comparing with GLCM based approach

Keywords: Content based image retrieval (CBIR), Haar wavelet, GLCM, Anatomical modality, Functional modality, Euclidean distance classifier.

1. INTRODUCTION

Medical imaging techniques were used to evaluate an area of body which are not extremely visible. Medical images were increasingly being used within healthcare and diagnosis, surgical planning, and treatment. The imaging modalities were divided in to two global categories: Anatomical and Functional imaging modalities. Anatomical modalities, i.e. depicting primarily the morphology, include X-RAY, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), US (Ultra Sound) [11]. Functional modalities i.e. depicting primarily information on the metabolism of the underlying anatomy, include SPECT (Single Photon Emission Tomography), PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging). CT/PET represents one of the medical imaging modalities with the largest growth worldwide. In 2009, approximately 2000 PET/CT scanners were installed in the United States and approximately 850 were installed in Europe. This large number of image collection shows increasing technical challenges to computer system to store/transmit, retrieve and index/manage image data effectively. Human intervention is very important in the CBIR process. Among visual features, texture is widely used for content-based access to medical images. The effectiveness of textural analysis depends on the methods used to extract meaningful feature.

In this proposed work, Section (ii) summarizes the methodology. Section (iii) describes the texture feature extraction, using GLCM. And Haar wavelet. Section (iv) describes the implementation for image retrieval Section (v) describes the performance evaluation. Section (vi) summarizes the results of performance evaluation.

2. METHODOLOGY

In this section the two texture based approaches were followed for medical image retrieval. The first approach by GLCM the second approach based on Haar wavelet based.

A. GLCM based texture feature extraction

This is one way of approach for texture based medical image retrieval. Fig .1. Shows the block of GLCM based retrieval.

B. Algorithm for GLCM based retrieval

- Read the query image.
- Compute the elements in the matrix
- Obtain energies.
- Repeat the above steps for databases.
- Find the Euclidean distance between database and query image.
- Sort the results in descending order.

c. Wavelet based texture feature extraction

The second approach using Haar wavelet based medical image retrieval. The steps followed by taking Discrete Wavelet Transform (DWT) and then applying Haar wavelet, then proceeded by computing total energy at sub bands and finding the energy level differences for dissimilarity computation. Fig .2. Shows the block of Haar wavelet based image retrieval.

D. Algorithm for Haar wavelet based retrieval:

- Read the query image.
- Decompose the image for six level decomposition.
- Obtain energies.

- Find the Euclidean distance between database and query image.
- Sort the results in descending order.

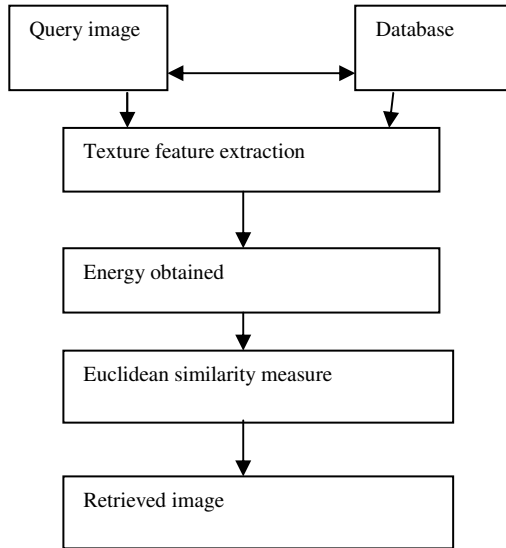


Fig . 1. GLCM based image retrieval

3. TEXTURE FEATURE EXTRACTION

Texture is one of the important features considered for image retrieval.

A. Texture Feature extraction using GLCM

GLCM is a statistical method for computing the co-occurrence probability of textural features. It expresses the texture features of gray-scale pixels at different positions.

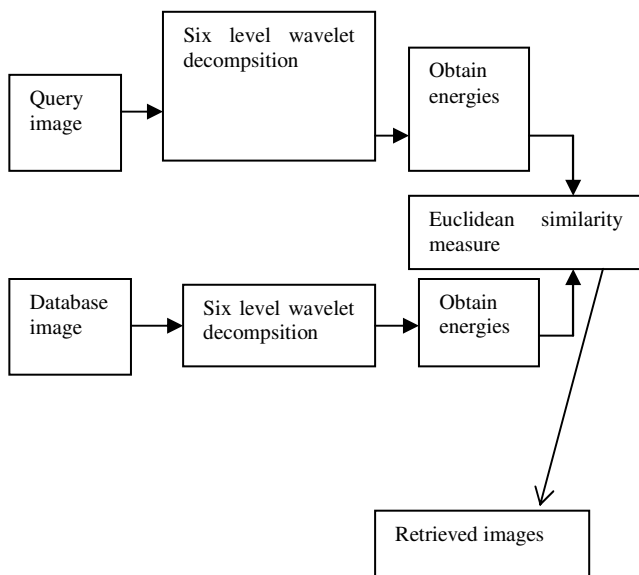


Fig . 2. Haar wavelet based image retrieval

The matrix $P(i,j,d,\theta)$ denotes the distance between pixels (x_1,y_1) and (x_2,y_2) . Elements in the matrix were computed by following equation.

$$P(i,j,d,\theta) = \frac{P(i,j,d,\theta)}{\sum_i \sum_j P(i,j,d,\theta)} \quad (1)$$

Texture features which can be extracted from gray-level co-occurrences matrix are as follows energy, entropy, contrast, inverse differences etc. Totally there are eleven texture features found. In this paper we have selected energy feature. $P(x,y)$ is the gray-level value at the coordinate (x,y) [1].

$$\text{Energy } E = \sum_x \sum_y P(x,y)^2. \quad (2)$$

Energy is a gray-scale image texture measure of homogeneity changing, reflecting the distribution of image gray-scale uniformity of weight and texture.

B. Texture Feature Extraction using DWT

DWT is used as a feature extraction and clarification tool, since its ability to localize structures with good resolution in a computationally effective manner. The wavelet transform utilizes both wavelet ϕ_r and scaling ϕ_k functions. The wavelet function is used to localize the high frequency content, whereas scaling function to examine low frequency. Haar transforms is the wavelet consisting of square shaped functions. It transforms signals from the space domain to a local frequency domain. A Haar wavelet decomposes an image using both low-pass filtering and high-pass filtering, working first on image columns and then on image rows. [8].

For a Haar wavelet

$$h_\phi = [h_\phi(0), h_\phi(1-0) = h_\phi(1)] = [1/\sqrt{2}, 1/\sqrt{2}] \quad (3)$$

Then,

$$h_\phi(0) = (-1)^0 h_\phi(1-0) = h_\phi(1) = 1/\sqrt{2} \quad (4)$$

$$h_\phi(1) = (-1)^1 h_\phi(1-1) = -h_\phi(0) = 1/\sqrt{2} \quad (5)$$

$$\phi(x) = \begin{cases} 1 & 0 \leq x \leq 1/2 \\ -1 & 1/2 < x < 1 \\ 0 & \text{elsewhere} \end{cases} \quad (6)$$

Scaling function $\phi(x)$ can be described as

$$\phi(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ \text{Otherwise} & \end{cases} \quad (7)$$

An average (approximation) component and a detail (fluctuation) component is transformed using Haar transforms. A signal with 2^n sample values, the first average sub signal $a^1(a_1, a_2, \dots, a_{n/2})$ for a signal length of N is given as follows

$$a = \frac{y_{2n-1} + y_{2n}}{\sqrt{2}}, n=1, 2, \dots, N/2 \quad (8)$$

And the first detail sub signal

$$d = (d_1, d_2, \dots, d_{N/2}) \text{ is given as} \quad (9)$$

$$d_n = \frac{y_{2n-1} - y_{2n}}{\sqrt{2}}, n=1, 2, \dots \quad (10)$$

The transform is applied to all rows of the matrix, the matrix by dimension (number of rows/2) \times (number of columns/2) as shown in the matrix. The matrix is classified as A the approximation area, H is the horizontal area, V is the vertical area and D is the diagonal area [8].

$$M = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix} \quad (11)$$

$$A = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad (12)$$

$$H = \begin{bmatrix} x_{13} & x_{14} \\ x_{23} & x_{24} \end{bmatrix} \quad (13)$$

$$V = \begin{bmatrix} x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \quad (14)$$

$$D = \begin{bmatrix} x_{33} & x_{34} \\ x_{43} & x_{44} \end{bmatrix} \quad (15)$$

4. IMPLEMENTATION & RESULTS OF PERFORMANCE EVALUATION

This technique can be implemented using MATLAB software. Similarity measure is to retrieve images. Here Euclidean similarity measure is implemented. CBIR system ranks similarity in descending order and then returns

relevant images that are most similar to the query images. The direct Euclidean distance between an image P and query image Q can be given as follows,

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^n (v_{pi} - v_{qi})^2} \quad (16)$$

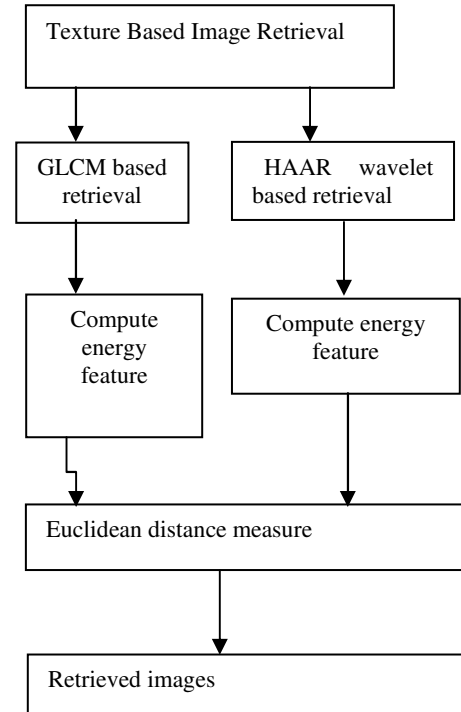


Fig. 3. Overall system of proposed work

The comparison of texture based image retrieval for GLCM based retrieval and wavelet based retrieval is performed. Fig 3. Shows the overall performance evaluation.

The medical images were collected and a database is formed. The A1 database consists of images like CT and MRI totally of 300 images. The F1 for PET image SPECT, FMRI of totally 300 images. They are first converted to gray scale. The algorithm was applied on these databases and their performances were evaluated.

Calculation of relevance is vital part in performance evaluation. Precision and Recall are basic measures used in evaluating the effectiveness of image retrieval system.

$$\text{Precision} = \frac{\text{Number_of_Relevant_Images_Retrieved}}{\text{Total_Number_of_Images_Retrieved}}$$

$$\text{Recall} = \frac{\text{Number_of_Relevant_Images_Retrieved}}{\text{Total_Number_of_Images_in_Database}}$$

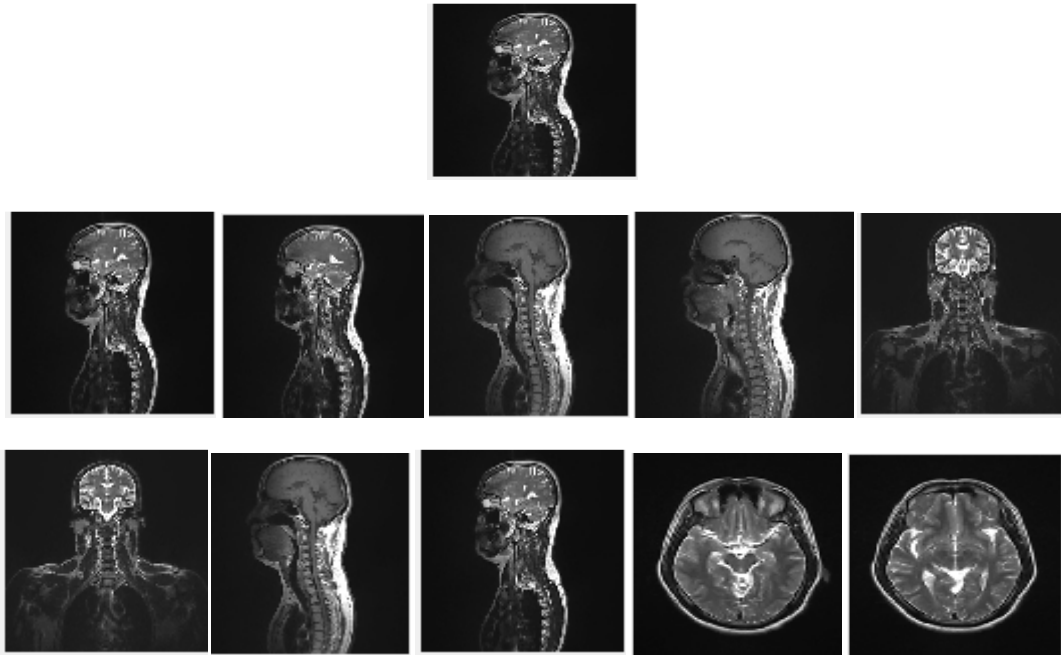


Fig. 4. First ten retrieved images using GLCM for anatomical modality

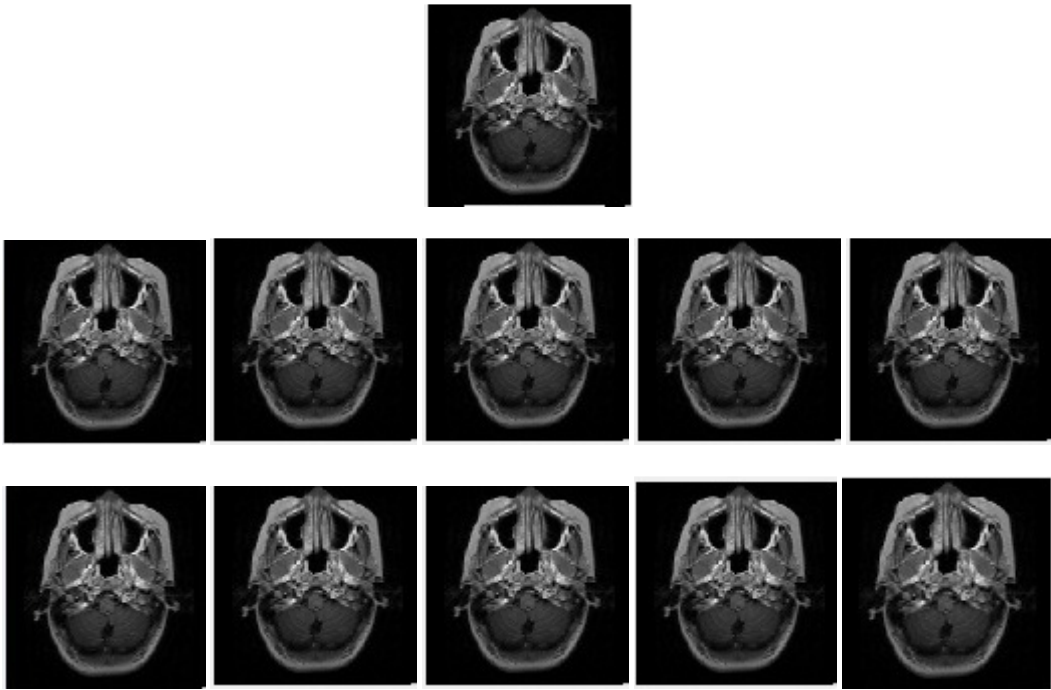


Fig. 5. First ten retrieved images using Haarwavelet for anatomical modality

The above described are to evaluate the results, to know the accuracy of the approaches. In this paper experimental dataset contain 600 medical images, divided into two categories they are anatomical and functional modalities. Each category has 300 images. Experimental images cover a

wealthy of content. Table 1 shows precision and recall results for anatomical modalities based on texture classification. Table 2 shows precision and recall results for functional modalities based on texture classification. Fig.4. Shows the first ten retrieved images using GLCM based

texture based image retrieval for anatomical modality. Fig.5. Shows the first ten retrieved images using Haar wavelet texture based image retrieval for anatomical image modality. Fig.6. Shows the first ten retrieved images using GLCM based texture based image retrieval for functional modality. Fig.7. Shows the first ten retrieved images using Haar wavelet texture based image retrieval for functional image modality.

Table 1 : Results From Retrieved Images

Retrieval mode	Precision(%)	Recall (%)	Number of relevant images retrieved
GLCM based retrieval	40	13	4
Haar wavelet based retrieval	80	30	8

Table 1i: Results From Retrieved Images

.Retrieval mode	Precision(%)	Recall (%)	Number of relevant images retrieved
GLCM based retrieval	30	11	3
Haar wavelet based retrieval	40	15	4

5. CONCLUSION

The main contribution of this work is to present texture based retrieval for both anatomical and functional modalities. The methodology is divided into image analysis and image retrieval stages. The purpose of image analysis is to collect samples from database and then apply it for feature extraction in the image retrieval stage. A CBIR system based on GLCM and Haar wavelet based was implemented. It was observed from the experimental results that, wavelet based medical image retrieval outperforms GLCM. The best precision ratio of 80 % and Recall ratio of 8 % were achieved using Haar wavelet. Rather than for functional modalities, anatomical modalities have achieved best relevant images.

REFERENCES

- [1] Fan-Hui Kong" Image Retrieval Using Both Color and Texture Features", in Proc of the Eighth Inte'l Conf. on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
- [2] Shang Lin, Yang YuBin, Wang Liang, Chen ZhaoQian, "An Image Texture Retrieval Algorithm Based on Color Co-
- occurrence Matrix (MCM)", Journal of NanJing University (Natural Science). Vol 40, No. 5, pp. 540-547, Sept.2004
- [3] H. T. Shen, B. C. Ooi, K. L. Tan, "Giving meanings to www images" Proceedings of ACM Multimedia, pp. 39-48, 2000.
- [4] B S Manjunath, W Y Ma, "Texture feature for browsing and retrieval of image data", IEEE Transaction on PAMI, Vol 18, No. 8, pp.837-842, 1996.
- [5] Y. Rui, C. Alfred, T. S. Huang, "Modified descriptor for shape representation, a practical approach", In: Proc of First Int's workshop on Image Database and Multimedia Search, 1996.
- [6] Cao LiHua, Liu Wei, and Li GuoHui, "Research and Implementation of an Image Retrieval Algorithm Based on Multiple Dominant Colors", Journal of Computer Research & Development, Vol 36, No. 1, pp.96-100,1999.
- [7] Song Mailing, Li Huan, "An Image Retrieval Technology Based on HSV Color Space", Computer Knowledge and Technology, No. 3,pp.200-201, 2007.
- [8] Kalyanasundaram.K, Sanoj.c.s "Content based Image retrieval using wavelet based multiresolution analysis",Int. Journal of Computer applications, Vol 40, No.4,pp 14-18, Feb 2012.
- [9] YANG Yubin, Chen Shifu, Lin Hui, "A Novel Image Retrieval Method Using Texture Features Based on Color-Connected Regions", ACTA ELECTRONICA SINICA, Vol 33, No. 1, pp. 57-62, Jan. 2005.
- [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] J.B. Antoine Maintz and Ma.A, Viergever, ASuevey of Medical Image Registration., Oxford University Press, Vol 2, 1998.
- [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):1267-1267, 2002.

Reconfigurability in Microstrip Patch Antennas

T. Sushma¹, N.V. Koteswara Rao², K. Rama Naidu³

¹Dept. of ECE, Methodist College of Engg. & Tech.

Osmania University, Hyderabad, India sushmatangella@yahoo.co.in

²Professor & Head, Dept. of ECE, Chaitanya Bharathi Instt. of Tech.,

Osmania University, Hyderabad, India, nvkoteswararao@gmail.com

³Professor, Dept. of ECE, JNTU, Pulivendula, A.P., India, kramanaidu@gmail.com

Abstract: Antennas are necessary and critical components of communication and radar systems. Different types of antennas have come up during the past few decades in both wireless communication and radar systems. The choice of an antenna imposes restriction on the overall system performance that arises because the antenna characteristics are fixed. Making antennas reconfigurable so that their behavior can adapt with changing system requirements or environmental conditions can eliminate these restrictions and provide additional levels of functionality for any system. This paper gives a brief review of characteristics of microstrip patch antennas and highlights the concept of reconfigurability in the context of antennas. Methods to achieve high gain and impedance bandwidth have been touched upon and emphasis is on related work that has been carried out in the area of reconfigurability in microstrip antennas in the past few years by several authors.

IndexTerms: High gain antenna, Impedance bandwidth, Microstrip patch antenna, Reconfigurability

1. INTRODUCTION

A microstrip antenna in its basic form consists of a metallic patch on a ground substrate, which is an extension of a microstrip transmission line. Advantages of microstrip antenna include low profile, conformable to planar and non-planar surfaces, inexpensive and simple to manufacture, and they are versatile in terms of resonant frequency, polarization, pattern and impedance. There are disadvantages of the microstrip antenna as well. They include low efficiency, high Q which results in a narrow frequency bandwidth, low power, poor scan performance and spurious feed radiation.

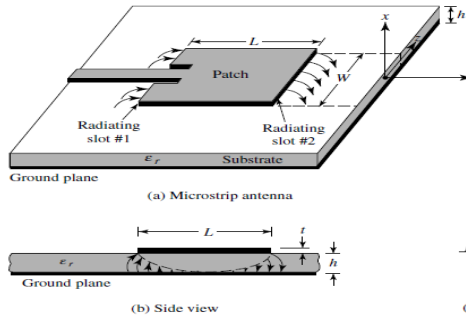


Fig. 1. Structure of a Microstrip Patch Antenna

Microstrip antennas [1] are often referred as patch antennas because the radiating element is normally a patch. This radiating patch comes in different shapes such as square, rectangular, circular, triangular and many other. There is a strong interest in the printed antennas for millimeter-wave applications. Due to the attractive features like ease of integration with the final RF stage of the communication system, many researchers around the world are attracted by the printed antenna technology.

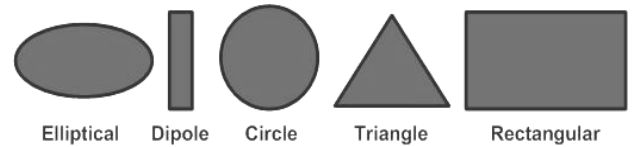


Fig. 2. Different shapes of patches

2. MICROSTRIP PATCH ANTENNAS

The foundation of printed antenna was originally started in 1953 when Deschamps proposed the use of microstrip feed lines to feed an array of printed antenna elements. Shortly thereafter, Lewin investigated radiation from strip line discontinuities. In the late 1960s Kaloj studied the basic rectangular and square configurations. However the microstrip element was first patented by Muson. Work done on basic rectangular and circular microstrip patches were published by Howell in the 1970s. Later Weinschel developed some microstrip geometries for use with cylindrical arrays. The application of microstrip element as an array was implemented by Sanford. Soon after the introduction of microstrip antenna, methods of analysis for these antennas were developed, including the transmission line model, the cavity model and the spectral-domain method.

Due to the several advantages of the microstrip antenna, many researchers [2, 3] have tried to overcome the disadvantages of a basic microstrip antenna as they can improve the performance of the antenna. Recently, various shapes of the microstrip antenna has been used such as slot patch antenna, H-shape, ring, triangle, U-shape and W-shape. Several feeding methods also are used such as coaxial

probe, transmission line, gap coupling and inset feed. Instead of using single element microstrip antenna, the arrays structure can be used to increase the gain of the antenna. Scaling factors technique and integration with active devices are also introduced in the microstrip antenna design. Based on the introduction of various shape of microstrip antenna, some parameters of the antenna can be changed such as to enhance bandwidth, increase gain, achieve circular polarization and improve the efficiency. The common shapes that are used for microstrip antenna design are rectangular, square, triangle, circle and hexagon. Shapes that are similar to letters can also be used for the microstrip antenna such as the letter O, H, F, W and E. The proximity and the aperture-coupled feed are two feeding techniques that can help in improving the bandwidth of the microstrip antenna.

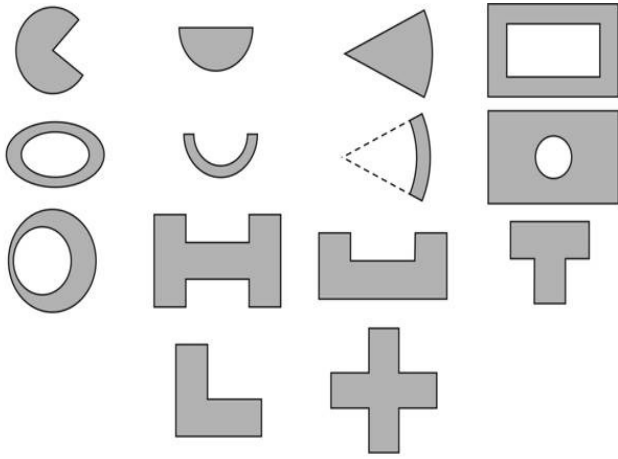


Fig. 3. Shapes of patches for special applications

Following is the summary of the effect of Microstrip Patch Antenna's (MPA) parameters:

- The antenna substrate dielectric constant is given as ϵ_r . Dielectric constant primarily affects the bandwidth and radiation efficiency of the antenna. Lower the permittivity, wider the impedance bandwidth and it also reduces the surface wave excitation.
- The antenna substrate thickness is given as h . The substrate thickness affects bandwidth and coupling level. A thicker substrate results in wider bandwidth, but less coupling for a given aperture size.
- L is the microstrip patch length. The length of the patch radiator determines the resonant frequency of the antenna.
- The microstrip patch width is given as w . The width, w of the patch affects the resonant resistance of the antenna; a wider patch gives a lower resistance.

3. RECONFIGURABILITY IN MICROSTRIP ANTENNAS

Reconfigurability [4], when used in the context of antennas, is the capacity to change an individual radiators fundamental operating characteristics through electrical, mechanical, or by other means. The traditional phasing of signals between elements in an array to achieve beam forming and beam steering does not make the antenna reconfigurable because the basic operating characteristics remain unchanged. Ideally, reconfigurable antennas should be able to alter their operating frequencies, impedance bandwidths, polarizations, and radiation patterns independently to accommodate changing operating requirements.

In many of the systems, there is a requirement to perform many functions across several frequency bands and operating bandwidths. In most cases, these requirements cannot be served by a single antenna but rather require the use of multiple antennas of varying factors and geometries. This results in an increase in fabrication costs, system weight, system volume, and resources required for maintenance or repair. Reconfigurable antennas can be used to address these system- requirements, given their ability to modify their geometry and behavior to adapt to changes in environmental conditions or system requirements (such as enhanced bandwidth, change in operating frequency, high gain requirement, polarization, radiation pattern etc.). Reconfigurable antennas can deliver the same throughput as a multi-antenna system using dynamically variable and adaptable single-antenna geometry. Reconfigurable antennas can thus provide great versatility in applications such as cognitive radio, MIMO systems, RFIDs, smart antennas, etc.

The advantages of using reconfigurable antenna compared to multi-band/wideband antennas or multiple antennas is its ability to support more than one wireless standard, minimizes cost, minimizes space requirement, allows easier integration, good isolation between different wireless standards, lower front-end complexity, capability to adapt and learn and in specific applications the antenna can be automated via a microcontroller or a field programmable gate array (FPGA).

The increase in demand for reconfigurable systems especially for wireless communications applications has stressed the need for smart RF devices that sense and respond to the RF changes in the environment. Many applications such as in a cognitive radio environment, antenna systems have to be designed to satisfy the reconfigurable multiple service and multi-band requirements which increase spectrum efficiency as well as the power utilization in modern wireless systems.

The different types of reconfigurable antennas are:

Frequency reconfigurable antenna: In this a radiating structure that is able to change its operating frequency by hopping between different frequency bands is used. This is achieved by producing tuning or notches in the antenna return loss.

Radiation pattern reconfigurable antenna: A radiating structure that is able to tune its radiation pattern is used for achieving reconfigurability. For this type, the antenna radiation pattern changes in terms of shape, direction or gain.

Polarization reconfigurable antenna [5]: A radiating structure that can change the polarization (horizontal/vertical, \pm slant 45° , Left-hand/Right-hand CP etc) is used. For this case the antenna can tune for example from vertical to left-hand circular polarization.

Multiple reconfigurable antennas [6]: Another type of antenna is a combination of the above three categories to exhibit many properties combined together. For example, one can achieve a frequency reconfigurable antenna with polarization diversity both at the same time. This finds application in smart array antennas.

Reconfigurable antennas have the capacity to change their operating characteristics through different methods that may be electrical, mechanical or optical. The basis for such antenna reconfiguration is redistribution of surface currents or of the effective aperture's electromagnetic fields. Such antennas have major benefits especially their ability to change their topology and behavior to adapt to environmental conditions and to remain synchronized with system requirements. The use of reconfigurable antennas in the past ten years has been based on traditional switching elements. Such switching elements (RF MEMS, PIN diodes, Varactors) require a complicated biasing networks that interfere with the antenna radiation mechanism as well as they add undesired resonances to the antenna performance.

The most common methodology adopted in the design of this type of antenna is based on the inclusion of some form of switches [7, 8]. Switches are used to connect and disconnect parts of an antenna or to simply redirect antenna surface currents across a slot. Antenna designer has to properly choose different types of switches as per their design requirements. The idea behind using a switch is to be able to control the flow of current across a certain structure. The ON state of a switch allows the current to pass through as in a short circuit; the OFF state prevents current from passing through. To control the activation and de-activation of a switch biasing networks are needed.

Recently optically gated photoconductive switches have come up. For such a switch to perform adequately and efficiently at microwave frequencies, it must have sufficient

voltage blocking capability which is necessary for adequate output power, and acceptable small on-state resistance and turn-on and turn-off times. Factors affecting these requirements include the properties of the semiconductor material itself, the electrical contacts, the dimensions of the switch, and the properties of the laser pulses used to gate the switch.

The recent advancements in this aspect leads to research and technology in merging of photonics and reconfigurable antennas in order to provide a new radio technology, which could be dynamically reconfigured at a very high speed. The design makes use of discrete semiconductor photoconductive elements, which can act as switches by dynamically changing their resonant frequencies. Regarding antenna design, the elimination of biasing circuitry would be a very important advantage in this approach, since bias lines can interfere with the operation of the antenna. The technology could be easily integrated into a fully reconfigurable wireless transceiver capable of enabling components with dynamically programmable parameters to improve the performance of current radio services such as wireless communication mobile services, medical sensing technologies, military radar and surveillance systems, and a variety of other radio applications.

Metamaterials are the materials whose permeability and permittivity are derived from their structure. These are artificial effectively homogenous electromagnetic structures with unusual properties not readily available in nature. The artificial structure materials are designed to interact with and control electromagnetic waves. Metamaterials include Right Handed Metamaterials (electromagnetic band gap structures (EBG) and left handed (LH) Metamaterials). EBG [9] are periodic arrangement of dielectric or metallic elements in one, two, or three dimensional manners. They prevent the passage of EM wave at certain angles of incidence at some frequencies.

Use of EBG results in reduced component size, reduces surface waves, increases radiation efficiency, increase in gain, reduces harmonics and reduces mutual coupling. Applications of EBG in the microwave domain is in developments concerning the direct control of the electromagnetic energy and its transmission, duplexers with controllable PBG materials [10], design low profile wire antennas with good radiation efficiency, etc.

4. HIGH GAIN ANTENNAS

High gain antennas [11] have many applications in wireless communication systems as they produce focused and narrow beamwidth, which allow for more precise targeting of the signal. Therefore, various gain enhancement techniques for antennas have been studied in the past decades. Specifically there are three important methods to achieve this purpose:

(1) Using antenna array [12], (2) Adding superstrate, and (3) Using intrinsically high gain antenna. For a very high gain, the array should contain a lot of elements and if the gain of each element is not high, it not only increases the size of the array, but also decreases the efficiency of it.

The benefit of using air as dielectric medium below inverted patch structure offers bandwidth increment. It also avoids penetration through the substrate avoided to accommodate the active devices and to have direct contact of co-axial probe with respect to patch and the ground plane. The application of using superstrate with inverted patch [13] offers gain enhancement and also it minimizes radiation losses as there is no necessity for drilling a hole through the patch. Use of parallel slots [14] also reduces size of the patch. On the other hand, use of superstrate provides the necessary protection for the patch from the environmental effects [15]. These techniques offer easy patch fabrication especially for antenna array structures.

5. WIDE BAND ANTENNAS

Modern wireless communication systems have significantly increased the demand for wideband antennas in order to support large number of users and to provide more information with high data speed [16]. The design of an efficient wide band small size antenna, for wireless applications, is a major challenge. Microstrip patch antennas have found extensive application in wireless communication system owing to their advantages such as low-profile, conformability, low-cost fabrication and ease of integration with feed networks. However, conventional microstrip patch antenna suffers from very narrow bandwidth, typically about 5% bandwidth with respect to the center frequency. This poses a design challenge for the microstrip antenna designer to meet the broadband techniques. There are numerous and well-known methods to increase the bandwidth of antennas, including increase of the substrate thickness, the use of a low dielectric substrate, the use of various impedance matching and feeding techniques [17], the use of multiple resonators, and the use of slot antenna [18], [19] geometry. However, the bandwidth and the size of an antenna are generally mutually conflicting properties, that is, improvement of one of the characteristics normally results in degradation of the other.

Recent research in the area of wideband antennas [20] includes a broadband gap coupled microstrip antenna using parasitic elements produce an impedance bandwidth, eight times that of a conventional patch antenna of the same size. By introducing a protrudent strip to a two-arm S-like monopole antenna, much wider impedance matching with multi-frequency resonant modes at the higher band is produced. To achieve maximum impedance bandwidth, a pair of notches can be placed at the two lower corners of the patch and the notch structure is embedded in the truncated

ground plane. To increase the impedance bandwidth of an antenna, a narrow slit can be used. By inserting an inverted U-slot [21] on the antenna, the frequency band notch characteristic is obtained.

A combination of beveling and shorting technique is used to increase the impedance bandwidth of the antenna. The new excitation structure consists of a planar monopole and a microstrip feed line, both of which are printed on the same dielectric substrate [22]. The wide-band performance is achieved by splitting the printed monopole with a slot.

Recently, several techniques have been proposed to enhance the bandwidth [23] which involves employing multilayer structures with parasitic patches of various geometries such as E, V and H shapes, which excite multiple resonant modes. However, these antennas are generally fabricated on thicker substrates.

6. CPW FED ANTENNAS

A Coplanar patch antenna [24] consists of a patch surrounded by closely spaced ground conductor on a dielectric substrate. CPW antennas [25] have many attractive features, such as no soldering points, easy fabrication and integration with monolithic microwave integrated circuits, and a simplified configuration with a single metallic layer. Thus, the designs of the CPW-fed antennas have recently received much attention.

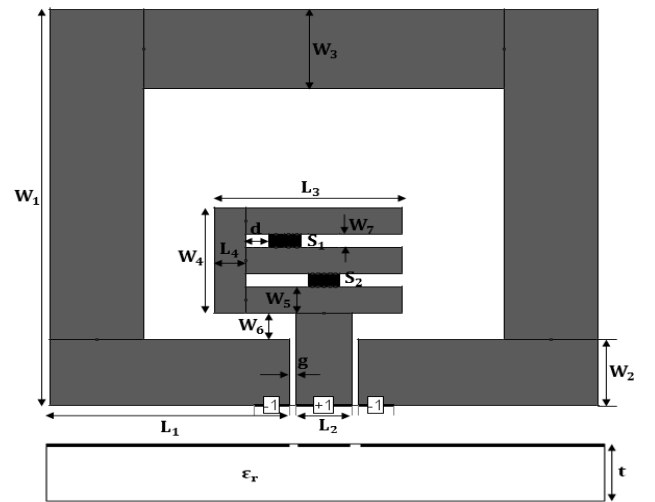


Fig. 4. A CPW-fed E-shaped reconfigurable antenna

7. RELATED WORK

Earlier, authors presented reconfigurable antennas, which radiated at different beam pattern by adjusting the apertures and maintaining their operating frequencies. For example a dual band dipole antenna integrated with MEMS switches [26] typically used a dual operating frequency to reconfigure

the beam pattern. Similarly, antennas were also proposed that worked at dual operating frequencies with a reconfigurable radiation pattern. In the reconfigurable antenna, the structure of the antenna can be changed by integrating with the switches such as PIN diode switches, the field-effect transistor (FET), photo conductor switches or by electromechanical system (MEMS) switch, which were proposed a few years ago. Works were presented that described and analyzed the reconfigurable corporate feed microstrip patch antenna incorporated with PIN diode as an RF switch. The switching mechanism is controlled by the external DC voltage. The antenna performances such as input return loss, bandwidth, half power beamwidth (HPBW), and radiation patterns obtained were appreciable.

In the recent years, multifunctional reconfigurable antenna (MRA) concept has gained significant interest. To dynamically change the properties of an MRA, the current distribution over the volume of the antenna needs to be changed, where each distribution corresponds to a different mode of operation. To this end, one can change the geometry of the antenna by switching on and off various geometrical metallic segments that make up the MRA. For switching, microelectromechanical system (MEMS), nanoelectromechanical system (NEMS) [27] or semiconductor type switches can be employed.

Today many wireless products are designed to operate at several frequencies. By making the antenna reconfigurable, one can accommodate more than one service using the same antenna and avoid the use of multiple antennas. In this case a reconfigurable antenna that maintains its radiation pattern at different frequencies is required. Recently reconfigurable stacked [28] Microstrip Patch Antenna (MSA) was proposed for wireless applications. The new antenna comprises of two layers; the bottom layer is a patch with two slots [29] designed on each side that can be controlled via switches. By adjusting the status of the switches, the resonance frequencies can be varied, thus achieving frequency reconfigurability. In order to increase the number of resonance frequencies and to enhance the bandwidth of the patch, another patch is placed on top of the first antenna. The two patches are separated with a dielectric layer optimized to yield the maximum number of resonance frequencies, bandwidth and gain. A polarization reconfigurable patch antenna [30] has also been designed that guarantees three different polarization states: a Right-Hand Circular polarization, a Left-Hand Circular Polarization and a Linear Polarization. Furthermore, frequency diversity has been also achieved for the Linear Polarization state.

8. CONCLUSION

The recent advancements in nanofabrication technologies have been opening up new applications in many disciplines including RF systems. The radiation and impedance analyses

of the NEMS integrated MRA are done. Currently, new nanofabrication processes for NEMS switch integrated reconfigurable antennas that can change not only frequency band but also polarization and radiation pattern of the radiated field is in progress.

The fundamental gain-bandwidth limitations of electrically small antennas prevent a small antenna from having high efficiency and wide bandwidth simultaneously. In the age of miniaturization, especially in the wireless communication industries, a promising solution to this limitation is to introduce reconfigurable antennas that can be tuned electronically to different frequency bands with both high efficiency and narrow instantaneous bandwidth.

In the above literature concentration was more on designing microstrip antennas with frequency reconfigurability that helps in enhancing the bandwidth. In view of the above, further research can be carried to enhance gain of a microstrip antenna/ arrays by incorporating reconfigurable techniques. These find applications in military/ defence systems. Microstrip antennas that guarantee multiple polarizations with reasonably high gain can also be designed with new materials like liquid crystals, meta materials, etc.

REFERENCES

- [1] David M.Pozar, "Microstrip Antennas", Proceedings of IEEE, Vol. 80, No.1, 1992.
- [2] Yong Liu, Li-Ming Si, Meng Wei, Pixian Yan, Pengfei Yang, Hongda Lu, Chao Zheng, Yong Yuan, Jinchao Mou, Xin Lv, and Housun Sun, "Some recent developments of Microstrip Antennas", International Journal of Antennas and propagation, Vol. 2012.
- [3] Chisang You, Manos M. Tentzeris and Woonbong Hwang, "Multilayer Effects on Microstrip Antennas for Their Integration with Mechanical Structures", IEEE Transactions on Antennas and Propagation, Vol.55, No.4, Pg:1051 – 1058, 2007
- [4] Zhang Jiajie, Wang Anguo, Wang Peng, "A Survey on Reconfigurable Antennas", ICMMT2008 Proceedings, June 2010
- [5] G. Monti, L. Corchia, and L. Tarricone, "Patch Antenna with reconfigurable polarization", Progress in Electromagnetics Research , Vol. 9, 13 -23, 2009
- [6] Symeon Nikolaou, Ramanan Bairavasubramanian, Cesar Lugo, Ileana Carrasquillo, Dane C. Thompson, George E. Ponchak, John Papapolymerou, Manos M. Tentzeris, "Pattern and Frequency Reconfigurable Annular Slot Antenna Using PIN Diodes", IEEE Transactions on antennas and Propagation, Vol. 54, No. 2, Pg.439-448, February 2006
- [7] J. Kiriazi, H. Ghali, H. Ragaie and H. Haddara, "Reconfigurable Dual-Band Dipole Antenna on Silicon using Series MEMS Switches", IEEE Antennas and Propagation Society International Symposium, 2003, vol.1 Pg:403 – 406, June 2003
- [8] W.B. Wei, Q.Z. Liu, Y.Z. Yin and H.J. Zhou, "Reconfigurable Microstrip patch Antenna with switchable polarization",

- Progress in Electromagnetics Research, PIER 75, 63-68, Pg.63 – 68, 2007
- [9] D.N. Elsheakh, H.A. Elsadek, E.A. Abdullah, H.M. Elhennawy and M.F. Iskander, "Ultra-Wide Bandwidth Microstrip Monopole Antenna by using Electromagnetic Band-Gap Structures", Progress in Electromagnetics Research Letters, Vol.23, Pg:109–118, 2011
- [10] X. Wang, M. Zhang and S.-J. Wang, "Practicability Analysis and Application of PBG Structures on Cylindrical Conformal Microstrip Antenna and Array", Progress in Electromagnetics Research, Vol.115, Pg:495 – 507, 2011
- [11] Mohammad Tariqul Islam, Mohammed Nazmus Shakib and Norbahiah Misran, "High Gain Microstrip Patch Antenna", European Journal of Scientific Research, ISSN1450-216X Vol.32, No.2, Pg:187 – 193, 2009
- [12] H. Attia, O. Siddiqui and O.M. Ramahi, "Artificial Magneto-superstrates for Gain and Efficiency Improvement of Microstrip Antenna Arrays", PIER Online, Vol.6, No.6, Pg:555 – 558, 2010
- [13] P.A. Ambresh, P.M. Hadalgi and P.V. Hunagund, "Study of Slot Inserted Inverted Patch—Rectangular Microstrip Antenna for Wireless Applications", International Journal of Electronics Engineering, 2 (2), Pg:295 – 298, 2010
- [14] Dimitrios Peroulis, Kamal Sarabandi, Linda P.B. Katehi, "Design of Reconfigurable Slot Antennas", IEEE Transactions on Antennas and Propagation, Vol.53, No.2, Pg: 645 – 654, February, 2005
- [15] A. A. Eldek, "Design of a High-Gain Cavity-Backed Slot Antenna With Mushroom Cells and Bent Ground Walls", Progress In Electromagnetics Research Letters, Vol. 20, 69-76, 2011
- [16] Mamdouh Gouda, Mohammed Y. M. Yousef, "Bandwidth Enhancement Techniques Comparison for Ultra Wideband Microstrip Antennas for Wireless Application", Journal of Theoretical and Applied Information Technology, Vol. 35 No.2, 31st January 2012
- [17] Jia-Yi Sze and Kin-Lu Wong, "Bandwidth Enhancement of Microstrip-Line-Fed Printed Wide-Slot Antenna", IEEE Transactions on Antennas and Propagation, Vol.49, No.7, Pg:1020–1024, 2001
- [18] Nader Behdad, Kamal Sarabandi, "A Varactor-Tuned Dual-Band Slot Antenna", IEEE Transactions on Antennas and Propagation, Vol.54, No.2, Pg:401 – 408, February, 2006
- [19] M. T. Islam, M. N. Shakib and N. Misran, "Multi-Slotted Microstrip Patch Antenna for Wireless Communication", Progress In Electromagnetics Research Letters, Vol. 10, 11-18, 2009
- [20] Nader Behdad, Kamal Sarabandi, "Dual-Band Reconfigurable Antenna with a very wide Tunability Range", IEEE Transactions on Antennas and Propagation, Vol.54, No.2, Pg: 409 – 416, February, 2006
- [21] G.F. Khodaei, J. Nourinia and C. Ghobadi, "A Practical miniaturized U-Slot Patch Antenna with enhanced Bandwidth", Progress in Electromagnetics Research B. Vol.3, Pg:47 – 62, 2008
- [22] Jacob George, C.K. Aanandan, P. Mohanan, K.G. Nair, H. Sreemoolanathan and M.T. Sebastian, "Dielectric-Resonator-Loaded Microstrip Antenna for Enhanced Impedance Bandwidth and Efficiency", Microwave and Optical Technology Letters, Vol.17, No.3, Pg:205 – 208, 1998
- [23] A.A. Abdelaziz, "Bandwidth Enhancement of Microstrip Antenna", Progress in Electromagnetics Research, PIER 63, Pg:311 – 317, 2006
- [24] M.A. Saed, "Reconfigurable Broadband Microstrip antenna fed by a Coplanar Waveguide", Progress in Electromagnetics Research, PIER 55, Pg:227-239, 2005
- [25] Ch. Sulakshana and L. Anjaneyulu, "A CPW fed E-shaped Reconfigurable Antenna with Frequency Diversity", International Journal of Information and Electronics Engineering, Vol.2, No.2, Pg. 174-177, March 2012
- [26] William H. Weedon, William J. Payne, Gariel M. Rebeiz, "MEMS-Switched Reconfigurable Antennas", IEEE Antennas and Propagation Society International Symposium, July, 2001
- [27] Bedri A. Cetiner, Necmi Biyikli, Bahadir S. Yildirim, Yasin Damgaci., "Nanoelectromechanical switches for Reconfigurable Antennas", Microwave and Optical Technology Letter / Vol. 52, No. 1, 64-69, January 2010
- [28] M.A. Alayesh, C.G. Christodoulou, M. Joler, S.E. Barbin., "Reconfigurable Multi-band stacked Microstrip Patch Antenna for Wireless Applications", Antennas and Propagation Conference, Pg. 329-332, 2008
- [29] K. Song, Y.-Z. Yin, B. Chen, S.-T. Fan and F. Gao, "Bandwidth Enhancement Design of Compact UWB Step-Slot Antenna with Rotated Patch", Progress in Electromagnetics Research Letters, Vol.22, Pg:39 – 45, 2011
- [30] M. Ramirez and J. Parron, "Circularly-Polarized stacked Annular-Ring Microstrip Antenna", Progress in Electromagnetics Research Letters, Vol.23, Pg:99 – 107, 2011

Medical Image Registration Based Retrieval Using Color and Texture Features

A. Swarnambiga¹, S. Vasuki², A. Ganesh Lakshmanan³

¹Ph.D Research Scholar, ²HOD & Professor, Velammal College of Engineering & Technology, Viraganoor, Madurai-09

³PG Student, Raja College of Engineering, Madurai

¹aswarnambiga@gmail.com, ²hello_vasuki@yahoo.co.in, ³aglakshman20@gmail.com

Abstract: This paper presents a new efficient algorithm for registration based retrieval of medical images using texture features for MRI images. This work focuses on retrieval using texture features in two categories. The first category is retrieval of MRI images using GLCM and the second category is retrieval using CCM for histopathological images. Then the new efficient registration based retrieval is processed. The color feature extraction is processed by HSV (Hue, saturation, Value) and by CCM (color co-occurrence matrix) for histopathological images. The texture feature extraction is obtained by using GLCM (Gray Level Co-occurrence Matrix) for MRI images. By quantifying GLCM, have combined Affine based registration with retrieval for MRI image. Image retrieval based on multi-feature fusion is achieved by using normalized Euclidean distance classifier. This new combined algorithm presents very good result in registration based retrieval of medical images.

IndexTerms: Image registration, Image retrieval, GLCM, CCM, CBMIR, CBIR, Euclidean Distance Method (EDM).

1. INTRODUCTION

Image retrieval is concerned with searching and retrieving digital images from a collection of databases. The need for efficient image retrieval has increased tremendously in many application areas such as biomedicine, military, commerce, education, and web image classification and searching. Effective searching for desired images from large-scale image database becomes an important and challenging research topic. Retrieval of image data based on pictorial queries is an interesting and challenging problem [1].

Image registration is the problem of finding a coordinate transformation that spatially aligns two or more images. It is a common necessity in applications of medical imaging. The images involved can be from different modalities, different time points, and/or different subjects. Image registration is an often encountered problem in many application areas like, for example, geophysics, medicine, and robotics. Here, we focus on medical applications. In the last two decades, computerized image registration has played an increasingly important role in medical imaging. Registered images are now used routinely in different applications, such as the treatment verification of pre- and post-intervention images and the

time evolution of an agent injection subject to patient motion. They are also useful to take full advantage of the complementary information coming from multimodal imagery, like, for example, computer tomography (CT) and magnetic resonance imaging (MRI). Given are two images, typically called reference R and template T. The goal is to find a spatial transformation, such that the deformed template matches the reference image subject to a suitable distance measure. In this example, we first compute an affine linear mapping such that the point wise difference, also called sum of squared differences, between the images.

This work focuses on the registration of pairs of images. One of the images, called the *moving* image, is deformed to fit the other image, the *fixed* image. The quality of alignment is defined by observer study and quantitative analysis, which measures the similarity of the fixed image and the deformed moving image. A high similarity leads to improved registration and vice versa.

The contents of an image have to be carefully extracted, classified, with different techniques for easy and efficient retrieval. The term 'content' in this context refers to colors, shapes, textures, or any other information that can be derived from the image itself. The retrieval on selected features usually yields images that have similar features. For example here Color and texture are the features used for retrieval. Larger the collection of images, greater is the chance that it contains an image similar to the query image. Retrieval by matching features of query image and database image is a recent trend.

In this work, a combined approach of using registration and retrieval is implemented. The first approach is individual registration using affine and individual retrieval using CCM and GLCM utilizing normalized Euclidean distance classifier for efficient medical image retrieval. The second approach is integrated registration based retrieval. The objective of this paper is to evaluate the use of color and texture as an image features for pattern retrieval [3, 4]. The proposed work in this paper is retrieval based on both texture and color feature extraction for individual and integrated approach. The remainder of this paper is

organized as follows: section (2) focuses on color based feature extraction, section (3) based on texture based feature extraction, Section (4) on registration based retrieval for medical images, section (5) on experimental result and evaluation, followed by discussion and conclusion.

2. COLOR BASED FEATURE EXTRACTION FOR HISTOPATHOLOGICAL IMAGES

A. Feature Extraction of HSV Color

In accordance with quantification, by quantified Hue (H), Saturation (S) and Value (V) a three dimensional feature vector with different weight to form an one-dimensional feature vector G [1][4]. The HSV space component used to reduce computation and improve efficiency. The user interface typically has query formulation part and result presentation part.

$$G = QSQVH + QV S + V \quad (1)$$

Where QS is quantified series of S , QV is quantified series of V . Here we set $QS = QV = 3$, then,

$$G = 9H + 3S + V \quad (2)$$

$$H = \begin{cases} 0 & \text{if } h \in [31, 20] \\ 1 & \text{if } h \in [21, 40] \\ 2 & \text{if } h \in [41, 75] \\ 3 & \text{if } h \in [76, 155] \\ 4 & \text{if } h \in [156, 190] \\ 5 & \text{if } h \in [191, 270] \\ 6 & \text{if } h \in [271, 295] \\ 7 & \text{if } h \in [296, 315] \end{cases} \quad (3)$$

$$S = \begin{cases} 0 & \text{if } s \in [0, 0.2] \\ 1 & \text{if } s \in [0.2, 0.7] \\ 2 & \text{if } s \in [0.7, 1] \end{cases} \quad (4)$$

$$V = \begin{cases} 0 & \text{if } v \in [0, 0.2] \\ 1 & \text{if } v \in [0.2, 0.7] \\ 2 & \text{if } v \in [0.7, 1] \end{cases} \quad (5)$$

In this way, three-component vector of HSV form one-dimensional vector, which quantize the whole color space for the 72 kinds of main colors. So can handle 72 bins of one-dimensional histogram. This is based on non interval quantization.

A. Color Feature Extraction based on CCM

Color image is divided into $N \times N$ image sub-block, for anyone image sub-block $T_{(i,j)}$ ($1 \leq i \leq N, 1 \leq j \leq N$), by main color image extraction algorithm to calculate the main color $C_{(i,j)}$. For any two 4-connected image sub-block $T_{(i,j)}$

and $T_{(k,l)}$ ($|i-k|=1$ and $j=l$; or $|j-l|=1$ and $i=k$). Image sub-block $T_{(i,j)}$ and $T_{(k,l)}$ are color connected. The entire image into a unique color of connected set $S = \{R_i\} (1 \leq i \leq M)$ in accordance with guidelines 4-connected region.

- 1) C_j and C_i belong to the same color of magnitude, that is, its HSV components $h_i = h_j, s_i = s_j, v_i = v_j$.
- 2) C_j and C_i don't belong to the same color of magnitude. It satisfy $s_i * 3 + v_i = s_j * 3 + v_j$, and $|h_i - h_j| = 1$; or satisfy $h_i = h_j, s_i = s_j, v_i, v_j \in \{0, 1\}$.

The statistic features extracted from CCM are as follows: Through this method, 8 dimensional texture features for component R, G in RGB color space and H in HSV color space is obtained.

Each component correspond to two statistic values E and S as Function $[RE, RS, GE, GS, HE, HS, VE, VS]$.

{For $E = \sum \sum$ for $(i=1,D), (j=1,D)$ }

$$Energy E = \sum \sum [m(i,j)]^2 \quad (6)$$

{For $I = \sum \sum (i,D), (j,D)$ }

$$Contrast I = \sum \sum (i-j)^2 \cdot m(i,j) \quad (7)$$

{For $S = \sum \sum (i,D), (j,D)$ }

$$Entropy S = \sum \sum m(i,j) \cdot \log[m(i,j)] \quad (8)$$

Where, if $m(i,j)=0; \log[m(i,j)]=0$

{For $H = \sum \sum (i,D), (j,D)$ }

$$Inverse\ difference\ H = \sum \sum \frac{m(i,j)}{1+(i-j)^2} \quad (9)$$

The necessity for additional, alternative access methods to the currently used, image based methods in medical information retrieval is detailed. It evaluates the need for image retrieval and presents concrete scenarios for promising future research directions.

3. TEXTURE BASED FEATURE EXTRACTION

A. Texture Feature Extraction Based on GLCM

GLCM creates a matrix with the directions and distances between pixels. Its level is determined by image gray level. GLCM is composed of the probability value, it is defined by $P_{\square}(i,j|d,\theta)$ which expresses the probability of the couple pixels \square at θ direction and d interval.

When θ and d is determined, $P(i, j|d, \theta)$ is showed by $P_{i,j}$. Elements in the matrix are computed by the following equation [1].

$$P(i, j|d, \theta)$$

$$P(i, j|d, \theta) = \frac{1}{\sum_i \sum_j P(i, j|d, \theta)} \quad (10)$$

GLCM texture feature is by the correlation of the couple pixels gray-level at different positions. Four features are selected, includes energy, contrast, entropy, inverse difference. Here $p(x, y)$ is the gray-level value at the coordinate (x, y) .

$$\text{Energy } E = \sum_x \sum_y p(x, y)^2 \quad (11)$$

{For $S = \sum (i, D), (j, D)$ }

$$\text{Entropy } S = \sum m(i, j). \log[m(i, j)] \quad (12)$$

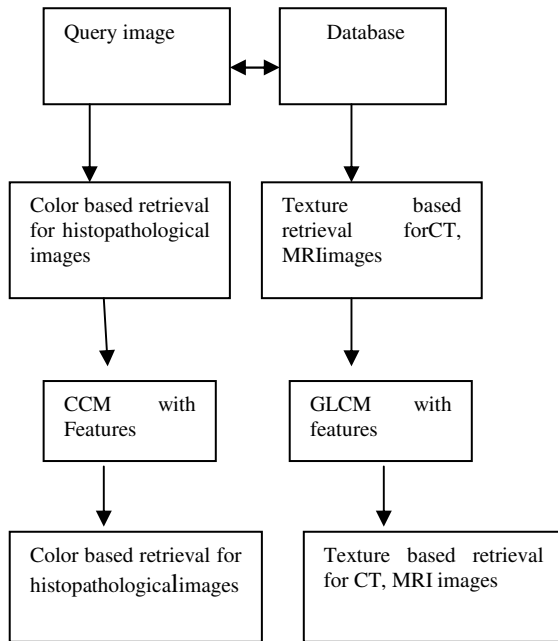


Fig. 1. Framework for retrieval of medical images using color and texture as a feature extraction techniques.

Retrieval of medical images was obtained by constructing the Euclidean calculation model.

$$D(A, B) = w_1 D(F_{CA}, F_{CB}) + w_2 D(F_{tA}, F_{tB}) \quad (13)$$

Normalized form as follows:

$$D(A, B) = w_1 \frac{\sqrt{2} - D(F_{CA}, F_{CB})}{\sqrt{2}} + w_2 \frac{\sqrt{2} - D(F_{tA}, F_{tB})}{\sqrt{2}} \quad (14)$$

4. AFFINE BASED REGISTRATION

We refer to global methods as the ones in which all pixels suffer the same transformation, which often results in simple and fast computation due to its small number of parameters. Rigid and Affine transformations are proposed as global transformations.

An affine transformation $T = \{[A], b\}$ in 2D space between point pair X and X_a is given by,

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (15)$$

It is a combination of several simple special mappings, such as the identity, translation, scaling, rotation, reflection and shear. Image acquisition is a projective process and if lens and sensor non linearities do not exist, the relation between two images of a rather flat scene can be described by projective transformation. When the scene is very far from camera, the projective transformation can be approximated by the affine transformation.

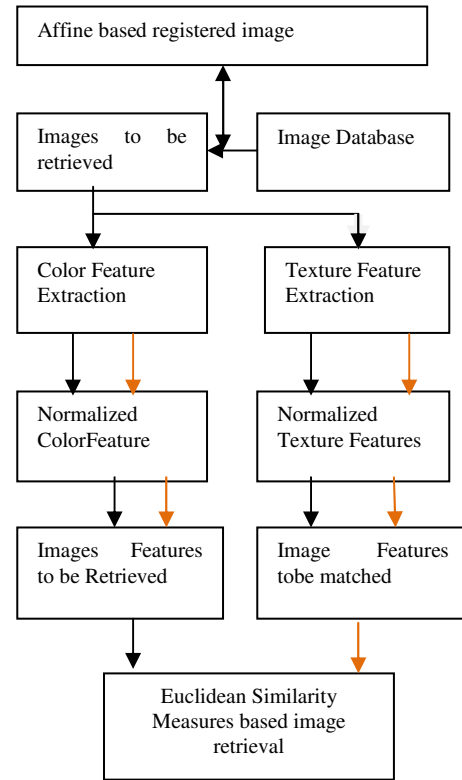


Fig. 2. The Framework for registration based retrieval

The main goal of this work is to review and evaluate the applicability of one-state-of-the-art intensity based image registration algorithm to brain images.

A) Similarity Metric

The measure of similarity between images or regions is a crucial component in image registration along with the selection of the transformation function. Using the similarity to drive the optimization process of the registration, similarity metrics are also often used to evaluate the performance of image registration. Higher similarity between images after registration means better alignment. Consequently, we will use similarity metrics as only one of the criteria for evaluating the most adequate registration method among those analyzed. In this section, we present the two metrics used in this work, sum of squared differences (SSD) and mutual information (MI).

1) SSD: The SSD metric computes the squared differences between intensity values for corresponding pixels. This is a simple, widely used metric that assumes a linear relationship between intensities in the images to be compared and its optimal value is 0 (images are identical). Where **A** and **B** stand for the images and iterates over the **I** pixels in the images, equation (15) shows how to compute this metric.

$$SSD(A, B) = \frac{1}{I} \sum_{i=1}^N (A_i - B_i)^2 \quad (15)$$

2) MI: MI provides a measure of probabilistic mutual dependence between two intensity distributions. MI allows to account for nonlinear differences in intensity (a feature often useful in multimodality registration) and is defined as

$$MI(A < B) = H(A) - H\left(\frac{B}{A}\right) = H(A) + H(B) - H(A, B) \quad (16)$$

Where **A** and **B** are the images to be compared

$$H(X, Y) = - \sum_{x,y=0}^N P(x, y) \log_2(p_{x,y}) \quad (17)$$

$$H(X) = - \sum_{i=0}^N p_i \log_2 p_i \quad (18)$$

Where (17) and (18) represents the joint and individual entropies, respectively, of random variables **X**, **Y** associated to the images to be compared. Here, **N** stands for the number of intensity levels and p_x (p_{xy}) is the probability value of **x** (**x,y**) in the (joint) probability distribution of variable **X** (**X** and **Y**).

B) Quantitative analysis experiment

From quantitative analysis experiment and metric evaluation it is found that, evaluating the results of registration methods in medical images is not an easy task. One could initially compute similarity metrics before and after registration to obtain an indication of how similar images are. A higher similarity is expected after image registration and the method with the highest similarity would be expected to be the most accurate. However, metric does not always tell the

full story as sometimes images that are “closer” in terms of metric functions are perceived to be more different by human observers. In order to analyze the correlation between similarity metric and visually correct registration, we also reviewed our methods using an observer study.

C) Observer study

In this part of the study, we aimed at evaluating the performance of registration methods using the subjective perception of experts. Experts were randomly presented with medical images registered using the different methods. For each image, they provided a subjective evaluation. Several criteria were used,

1) Registration artifacts (ARTIF): For instance, features those were not present in the template image but are present in the registered image or unrealistic deformations.

2) Visual similarity (RESULT): Between registered template and target images using the experts.

3) Difference image (DIFF): This allows evaluating dissimilarities and registration performance from a global point of view.

In our case (ARTIF) and difference image is found to be less which can be avoided and carried for next step. Whereas the experts view by their visual observance and they also confirms the similarity which is found to be closer. Hence this observer study also confirms effective registration.

5. EXPERIMENTAL RESULTS AND EVALUATION

For retrieval efficiency, traditional measures namely Precision and Recall were computed with 500 medical images as test samples. Standard formulas have been computed for determining the Precision and Recall measures. For the analytical notion of performance along with the subjective evaluation, We used the traditional precision-recall value (PR) performance metrics measured under relevant (and unbiased) conditions. The precision and recall values are given as

$$R = \frac{RR}{TR} \quad (19)$$

$$P = \frac{RR}{N} \quad (20)$$

Where **RR** is the number of relevant items retrieved (i.e., correct matches) among total number of relevant items, **TR**. **N** is the total number of items (relevant + irrelevant) retrieved. By randomly selecting some sample query images from the MATLAB – image processing tool box-

Workspace database, the system was tested and the results are shown. The query processing and feature extraction with distance classification are the two main processes in retrieval of images. The Euclidean distance is used for distance classification and energy as feature extraction technique in the retrieval of medical images. The Euclidean distance metrics and energy as the measure is followed as second step in the retrieval of medical images. The direct Euclidean distance calculation between a database image p and query image q is given below. By using equation (21) euclidean distance metric is made possible. Where $i=1$ to n .

A) Registration based Retrieval Strategies

To our knowledge this is the first attempt to quantitatively combine registration and retrieval for medical images. Specifically taking in to account the quantitative and

evaluation criteria used in this work: metric comparison, an observer study, running time analysis were evaluated for a subset of images before and after registration. For registration based retrieval efficiency, precision and recall values were calculated. Subsequently, registration results are presented, providing details on the data, experiments, and quantitative analysis for registration based retrieval for medical images.

$$\text{Error rate} = \frac{\text{Number of non relevant images retrieved}}{\text{Total number of images retrieved}} \quad (21)$$

$$\text{Retrieval efficiency} = \begin{cases} \frac{\text{Number of relevant iamges retrieved}}{\text{Total number of images retrieved}} \\ \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \end{cases} \text{ If number of retrieval} > \text{number of relevant or Else} \quad (2)$$

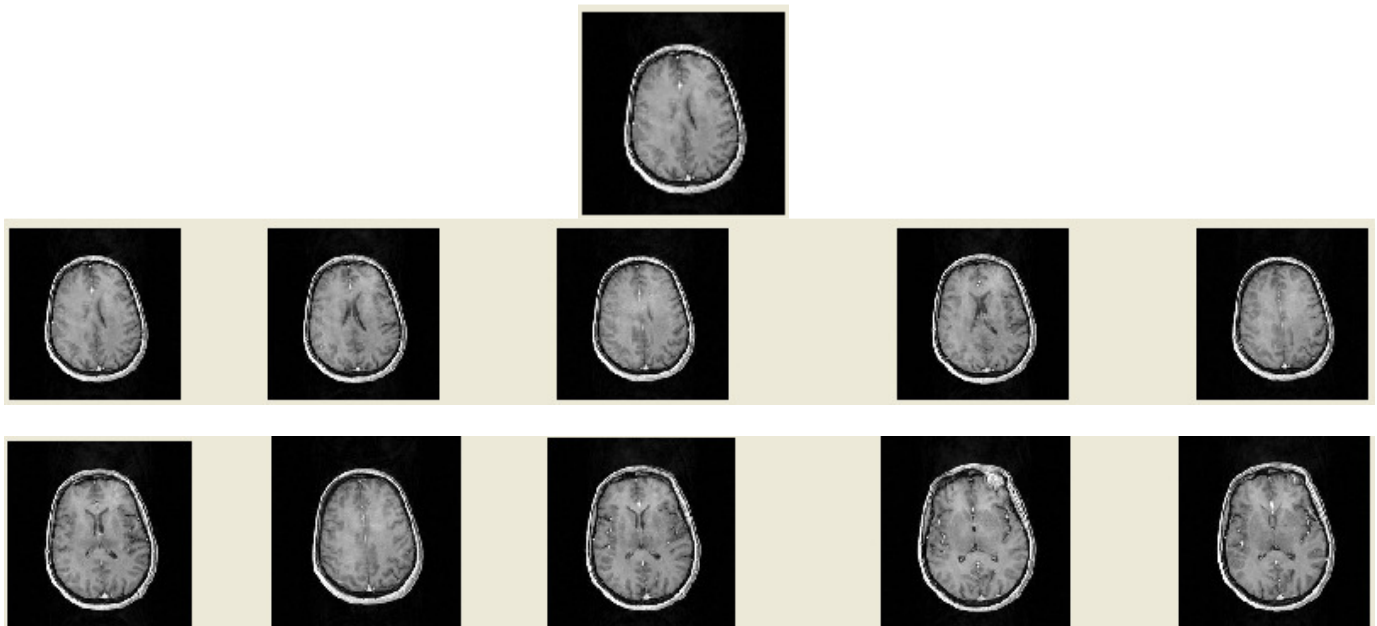
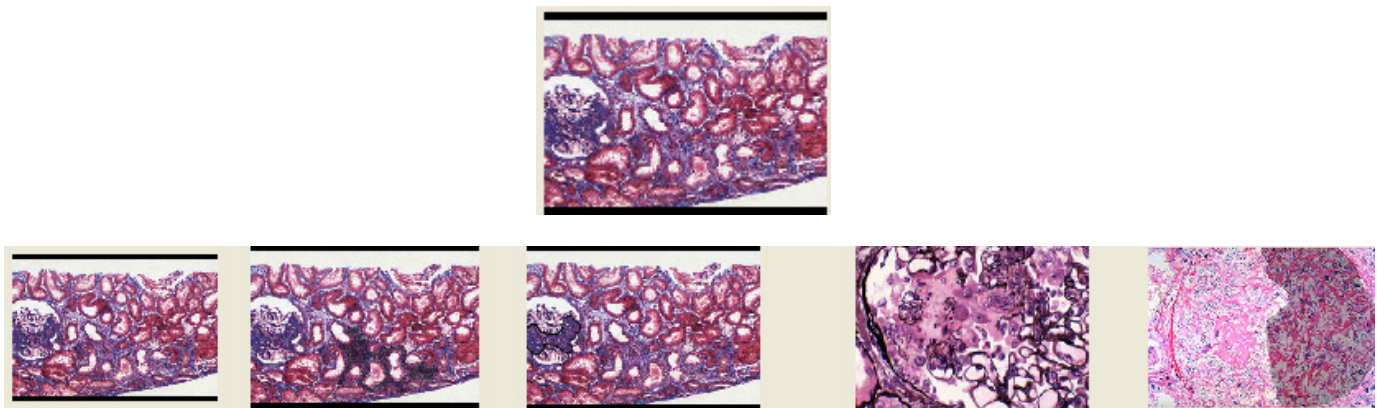


Fig. 3 . The retrieved images using GLCM and top left is the query image.



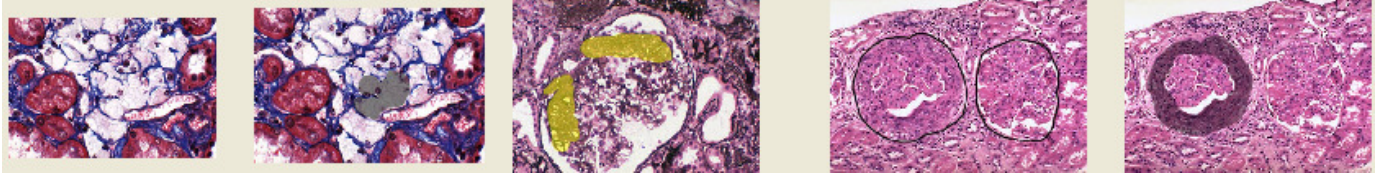


Fig. 4. The retrieved images using CCM and top left is the query image.

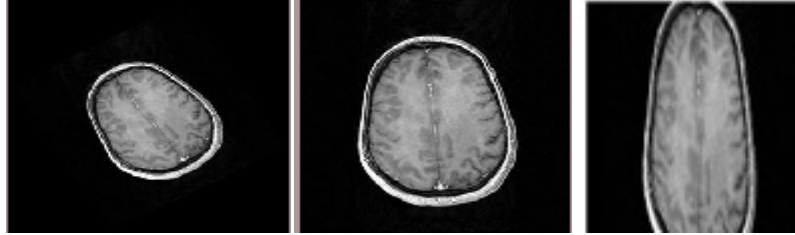


Fig. 5 . Top left is the test image to be registered and middle is the reference image and top right is the registered image.

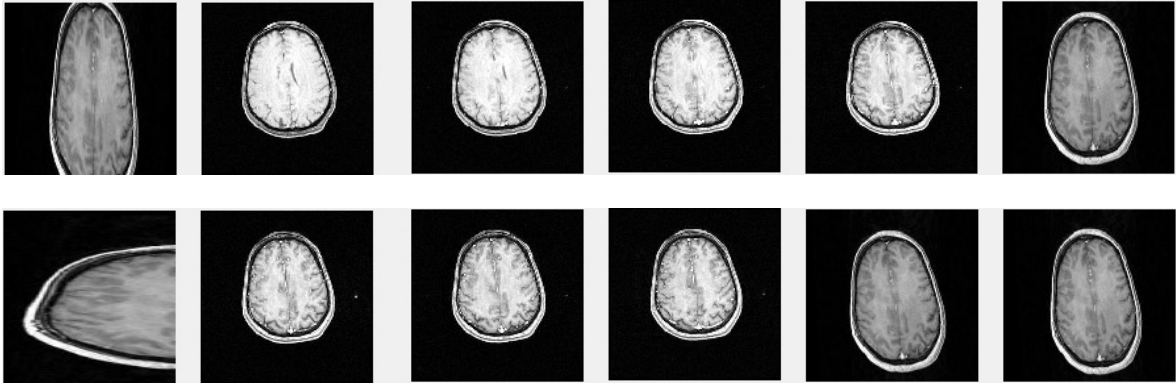


Fig. 6. Affine based registration with GLCM based retrieved image.

Table I: Performance calculation for retrieval without registration

Performance calculation	GLCM	CCM
Number of relevant retrieved images	10	4
Number of irrelevant retrieved images	NIL	6
Number of actual images retrieved	1	4
Number of relevant images in the database	100	100
Precision	70	40
Accuracy	50	20
Efficiency	70	50
Error Rate	18	30

Table II: Performance calculation for retrieval with registration

Performance calculation	GLCM
Number of relevant retrieved images	10
Number of irrelevant retrieved images	NIL
Number of actual images retrieved	1
Number of relevant images in the database	1
Precision	90
Accuracy	60
Efficiency	80
Error Rate	8

B) Discussion

From the experiment it is clear that Registration based retrieval is performing good in retrieval with a precision of 80% , accuracy of 60% and with a lesser error rate showing an efficiency of 80%.

Where as this registration based retrieval is shown for only gray scale medical image MRI. Registration is applicable only to anatomical and functional modalities, not for pathological images. But our experimental setup is tried for texture based image retrieval in two categories. The first category using GLCM based texture based image retrieval for MRI image. The retrieved image is shown in the Fig. 3. Then the second category for texture based image analysis using CCM is experimented with histopathological images. The retrieved image is shown in the Fig. 4. Then the affine based image registration is carried with MRI images which is applied to Affine based registration with GLCM for MRI brain retrieval . The affine based registration is shown in Fig. 5. The registration based retrieved images is shown in Fig. 6.

Fig. 6. The experimental setup is tried with 30 training sets and the retrieved results are given in table I and Table II.

6. CONCLUSION

Thus it is proved that registartion based retrieval performs better than GLCM based retrieval even though with their various geometrical orientation. The histopathological images also do perform good on Retrieval. The pathological based retrieval is found to be difficult , but our experimental setup do showed a good result.. This work hence proved the results for texture based retyrieval results for MRI image and histopathological image. And also proved registartion based retrieval for MRI iamges. This will be helpful for Clinical diagnosis, treatement planning and surgical guidance.

The work can be further extended to Extrinsic registartion based retrieval and Intrinsic registartion based retrieval which will be helpful in clinical diagnosis..

REFERENCES

- [1] Fan-Hui Kong” Image Retrieval Using Both Color and Texture Features”,in Proc of the Eighth Inte’l Conf. on Machine Learning and Cybernetics, Baoding, 12-15 July 2009
- [2] H. T. Shen, B. C. Ooi, K. L. Tan, “Giving meanings to www images” Proceedings of ACM Multimedia, pp. 39-48, 2000.
- [3] B S Manjunath, W Y Ma, “Texture feature for browsing and retrieval of image data”, IEEETransaction on PAMI, Vol 18, No. 8, pp.837-842, 1996.
- [4] Y. Rui, C. Alfred, T. S. Huang, “Modified descriptor for shape representation, a practical approach”, In: Proc of First Int's workshop on Image Database and Multimedia Search, 1996.
- [5] Cao LiHua, Liu Wei, and Li GuoHui, “Research and Implementation of an Image Retrieval Algorithm Based on Multiple Dominant Colors”, Journal of Computer Research & Development, Vol 36, No. 1, pp.96-100, 1999.
- [6] J. R. Smith, F. S. Chang, “Tools and Techniques for Color Image Retrieval”, Symposium on Electronic Imaging: Science and Technology-Storage and Retrieval for Image and Video Database IV, pp.426-237, 1996.
- [7] Song Mailing, Li Huan, “An Image RetrievalTechnology Based on HSV Color Space”, Computer Knowledge and Technology, No. 3, pp.200-201, 2007.
- [8] Shang Lin, Yang YuBin, Wang Liang, Chen ZhaoQian, “An Image Texture Retrieval Algorithm Based on Color Co-occurrence Matrix (MCM)”, Journal of Nan Jing University (Natrual Science). Vol 40, No. 5, pp. 540-547, Sept.2004.
- [9] YANG Yubin, Chen Shifu, Lin Hui, “A Novel Image Retrieval Method Using Texture Features Based on Color-Connected Regions”, ACTA ELECTRONICA SINICA, Vol 33, No. 1, pp. 57-62, Jan. 2005.
- [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 onComputer and Robot Vision*, 2005.
- [12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image Segmentation UsingExpectation-maximization and Its Application to Image Querying,” *IEEE Trans. Pattern Analysis and MachineIntelligence*, 24(8):1026-1038, 2002.
- [13] B. Fischer, J. Modersitzki, Curvature based image registration, *JMIV* 18 (1) (2003) 81–85.
- [14] B. Fischer, J. Modersitzki, Combination of automatic non-rigid and landmark based registration: the best of both worlds, in:M. Sonka, J.M. Fitzpatrick (Eds.),*Medical Imaging 2003: Image Processing*, Proc. SPIE, vol. 5032, 2003, pp. 1037–1048.
- [15] B. Fischer, J. Modersitzki, FLIRT: A Flexible Image Registration Toolbox, in: J.C. Gee, J.B.A. Maintz, M.W. Vannier (Eds.), *Biomedical Image Registration, Second International Workshop, WBIR 2003, LCNS*, vol. 2717, Springer, Berlin, 2003, pp. 261–270.
- [16] S. Haker, A. Tannenbaum, R. Kikinis, Mass preserving mappings and image registration, in: *MICCAI 2001, LNCS*, vol. 2208, 2001, pp. 120–127.
- [17] S. Henn, K. Witsch, A multigrid approach for minimizing a nonlinear functional for digital image matching, *Computing* 64 (4) (1999) 339–348.
- [18] B.K.P. Horn, B.G. Schunck, Determining optical flow, *Artificial Intell.* 17 (1981) 185–204.
- [19] M. Lefébure, L.D. Cohen, Image registration, optical flow and local rigidity, *JMIV* 14 (2) (2001) 131–147.
- [20] J.B.A. Maintz, M.A. Viergever, A survey of medical image registration, *Medical Image Anal.* 2 (1) (1998) 1–36.
- [21] C.R. Maurer, J.M. Fitzpatrick, Interactive image-guided neurosurgery, in: *A Review of Medical Image Registration*, American Association of Neurological Surgeons, Park Ridge, IL, 1993, pp. 17–44.
- [22] M.I. Miller, L. Younes, Group actions, homeomorphisms, and matching: a general framework, *Int. J. Comput. Vision* 41 (1/2) (2001) 61–84.
- [23] D. Potts, G. Steidl, Optimal trigonometric preconditioners for nonsymmetricToeplitz systems, *Linear Algebra Appl.* 281 (1998) 265–292.
- [24] A. Roche, Recalaged’ images médicales par inférencestatistique, Ph.D. Thesis, Université de Nice, Sophia-Antipolis, France, 2001.

A Multi Band Switchable Circularly Polarized Slotted Microstrip Patch Antenna

Ajit Yadav¹, Shweta Gautam², Mithilesh Kumar³

^{1,2,3}Electronics Department, University College of Engineering, Rajasthan Technical University, Kota
¹ajity2@gmail.com, ²s.ecom.og@gmail.com, ³mith_kr@yahoo.com

Abstract: The design of multi band and switchable circularly polarized Slotted Microstrip patch antenna is presented. The antenna is excited by Microstrip feed line. The proposed antenna has square patch with a ring slot in the square patch, and the circular polarization radiation can be generated by current perturbation due to diodes connected across the ring slot to the circular patch. The p-i-n diodes across the ring slot alter the direction of current path. Three p-i-n diodes simply switch the direction of circular polarization between the left handed CP and right handed CP. Linear polarization can also achieved if both side ways diodes are OFF and centre diode is ON. Key parameters have been analysed on Electromagnetic simulation software. The operating frequencies are 1.55GHz, 3.8GHz and 4.77 GHz and results obtained are satisfactory.

Index Terms: Circular polarization (CP), microstrip slotted patch antenna, polarization switching, left hand circular polarization (LHCP), right hand circular polarization (RHCP)

1. INTRODUCTION

Aim of antenna design is to transmit large amount of information to the longer distance, so the research in wireless communication is growing toward the reduction of antenna size and get the higher transmission bandwidth of antenna. Many applications require integrated multifunctional terminals [1], [2]. The concept of frequency reconfigurable and polarization reconfigurable is developed. The pursuit of antenna banks with compact form factors has led to research into reconfigurable antennas with variable bandwidth [3], radiation characteristics [4], polarization [5] and for achieving the multiple bands cascaded antennas are designed. In the cascaded antenna large number of antennas are connected to form a single antenna and the switching is done to get the required frequency band. Such mobile terminals must operate while acting as antenna banks because each service requires a different frequency band. This structure increases the information rate but one problem is there, that is size of antenna. Accomodating large size of antenna in small devices is not possible, so the research is done in area of reducing the size of antenna.

Polarization diversity with switchable frequency is also used in order to reduce fading in wireless local area networks (WLANs) [6], as a modulation scheme in radio frequency

identification (RFID) systems [7], and to increase the security complexity in military wireless. The dual band antenna is simply obtained by connecting the square and circular patch. But still there is a need of higher data information transmission rate, for this purpose polarization switching concept is introduced. In polarization switching large amount of information can be transmitted through a single feed port without interference. This is achieved by orthogonal polarization modes that are left hand circular polarization (LHCP) and right hand circular polarization (RHCP). A switchable CP slot antenna with a p-i-n diode switch for different polarizations has also been reported [8]. It is also possible to achieve switchable CP microstrip antennas with shorting walls that are switched to the ground plane via p-i-n diodes [9], [10]. But by connecting the two patches at one more position will generate a new band without increasing the size of antenna.

In this paper a multi band polarization switchable antenna has been proposed. The structure is single feed simple and compact producing multi band, circular and linear polarization. The orthogonal modes for circular polarization are generated by p-i-n diodes connecting rectangular and circular patch. The polarization switching is obtained by switching the p-i-n diodes. Different polarization has been discussed, which are obtained by switching the diodes. In subsequent sections antenna structure and results have been discussed. Section IV provides conclusion and possible application of proposed antenna.

2. ANTENNA STRUCTURE

Fig.1 shows the geometry of the proposed multiband and dual switchable circular polarization microstrip slotted patch antenna. Proposed antenna has three layer structures, bottom layer is a ground plane of thickness .01mm, on which there is a second layer of dielectric substrate FR4 of thickness 2 mm and dielectric constant is 4.9. Third layer is a patch which is square shaped with a ring slot. The square patch length is L mm. At the centre of the square patch there is a ring slot of outer radius R2 and the inner radius R1. Fig.1 shows the construction of the square patch with a ring slot inside it. This ring slot separates the two structures one is square patch and other circular patch which is inside the

square patch. The separation between the square patch and circular patch which is inside it is 2mm or the width of ring slot is 2mm. This proposed antenna produces linear polarization as well as circular polarization by introducing diode.

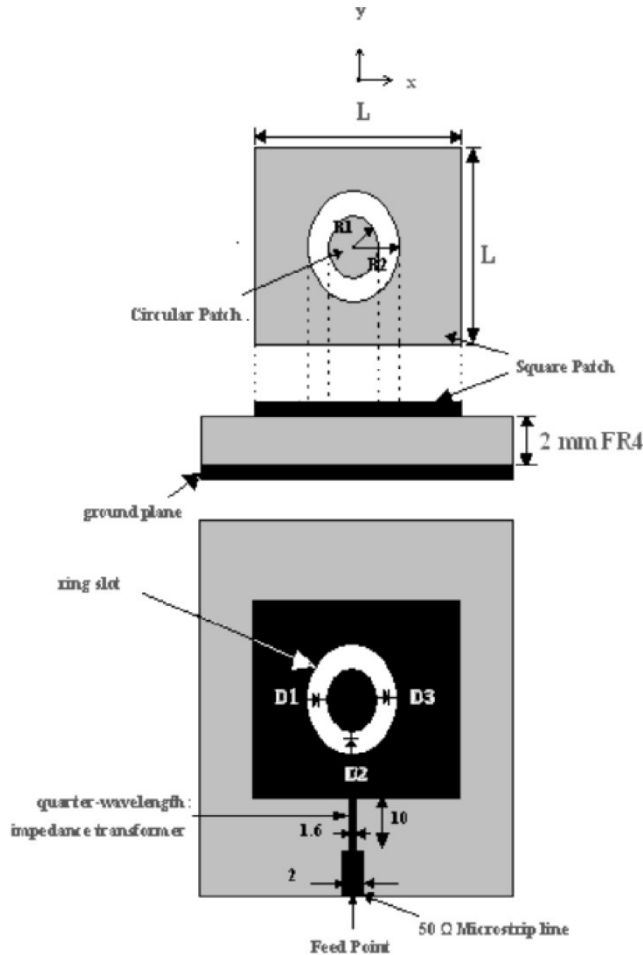


Fig.1. Geometry of multiband and dual switchable circular polarization microstrip slotted patch antenna Dimensions: 80*80 mm²

Fig.1 shows p-i-n diodes D1, D2 and D3. These p-i-n diodes providing the connection between the square patch and the circular patch. The diode D1 is connected between square patch and inner circular patch in left side, and diode D3 is connected between square patch and inner circular patch in right side and diode D2 is at down side. The polarization diversity is obtained by switching different p-i-n diodes at different time. Linear polarization is obtained when P-I-N diode D2 is forward biased and D1 and D3 are reversed biased. To get circular polarization two currents are fed at 90° generating orthogonal components. The LHCP is obtained when left p-i-n diode i.e. D1 is forward biased along with D2, because of current direction. Similarly RHCP is obtained when the right p-i-n diode (D3) is forward biased

along with D2, reverse biasing the D1. The antenna structure is three bands polarization switching is obtained in only one band which is generated by circular patch. It is seen that axial ratio varied from 40 dB to 1 dB.

Table I: Summary of Antenna Polarization

P-I-N diodes Biasing	Polarization
D2 : ON (Forward biased) D1,D3: OFF (Reverse biased)	Linear Polarization (LP)
D2,D3: ON (Forward Biased) D1 : OFF (Reverse Biased)	Right Hand Circular Polarization (RHCP)
D1,D2: ON (Forward Biased) D3 : OFF (Reverse biased)	Left Hand Circular Polarization (LHCP)

3. IPLIMENTATION AND RESULTS

The proposed antenna is designed for multibands (1.55GHz, 3.8GHz and 4.77 GHz). The square patch ($L=40\text{mm}$) on the top of the FR4 substrate with $\epsilon_r=4.4$ and height $h=2\text{mm}$. one ring slot in the square patch with dimensions outer radius ($R2=7\text{ mm}$) and inner radius ($R1=5\text{ mm}$), width of ring slot is 2mm, the three p-i-n diodes are located across the slot in order to control the current path. The $\lambda/4$ impedance transformer is used for impedance matching between patch and 50Ω feed line.

TABLE I shows the summary of the antenna polarization. With the above said dimensions the resonate frequencies for the linear polarization (when diode D2 is forward biased and remaining two are reversed biased) are 2.1 GHz and 5.1 GHz. The band generated by the circular patch can be switched to the circular polarization by switching on either D1 or D3 we will get three bands at this point. The bands generated by circular patch can be RHCP or LHCP according to the position of p-i-n diode forward biased.

Fig.2 shows the simulated results of proposed antenna when diode D1 and diode D3 are OFF and only diode D2 is ON, in this case antenna radiate at two frequency bands at 2.1GHz

and at 5.1GHz. In this case the proposed antenna is linearly polarized.

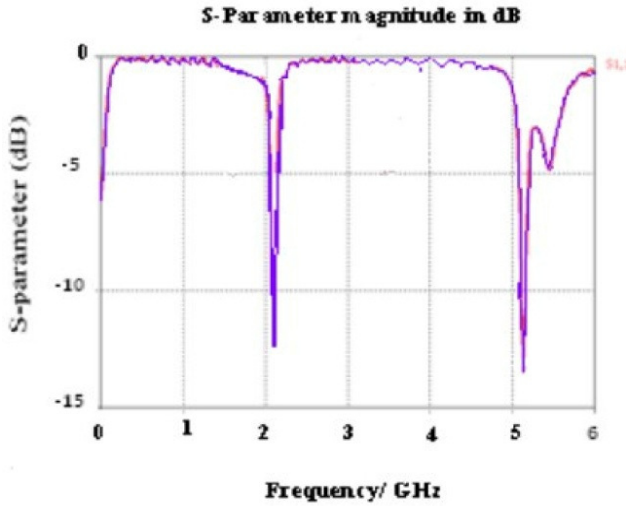


Fig. 2. Simulated s-parameter of the proposed antenna when (D1=OFF, D2=ON, D3=OFF)

Fig.3 shows the simulated results of proposed antenna at right hand circular polarized mode and left hand circular polarized mode. In LHCP case diode D1 and diode D2 are in ON state and diode D3 is in OFF state, there are three frequency bands at which the antenna radiates

Fig.4 shows the simulated farfield pattern of proposed antenna. In RHCP three farfield patterns are observed at three different frequencies. At frequency 1.55GHz the main lobe magnitude is 6.2dBi and the main lobe direction is 0.0 deg. At frequency 3.8GHz the main lobe magnitude is 3.7 dBi and the main lobe direction is 30.0 deg. And at 4.77GHz frequency the main lobe magnitude is 7.6 dBi and the main lobe direction is 5.0 deg.

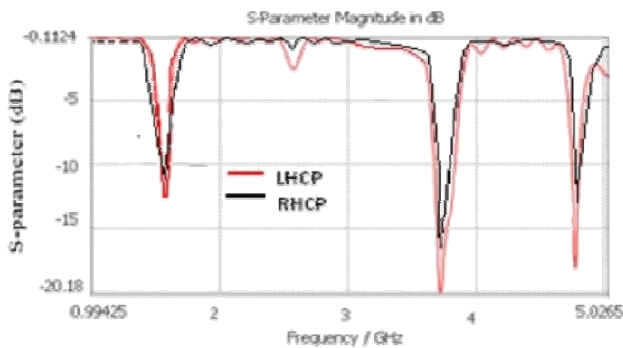


Fig.3 simulated s-parameter of proposed antenna when (D1=ON, D2=ON, D3=OFF (LHCP)) and when (D1=OFF, D2=ON, D3=ON (RHCP)) (a) (b)

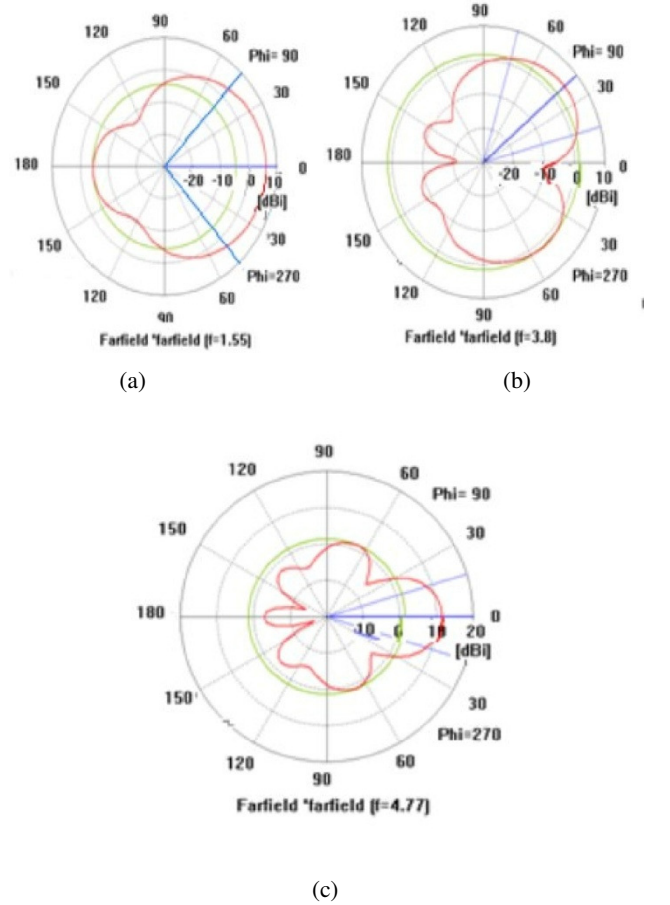
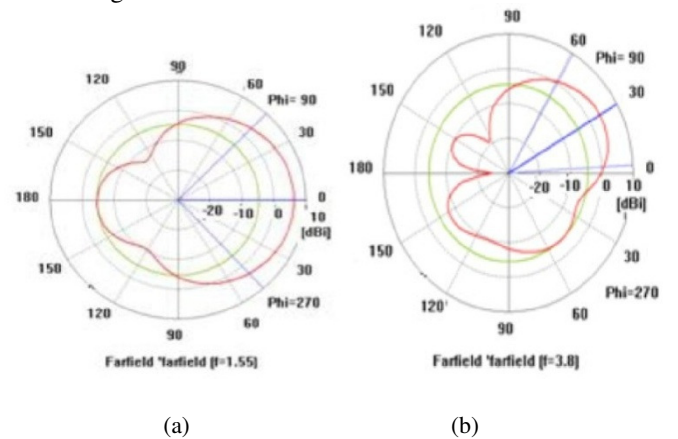


Fig. 4. Simulated Farfield Pattern of RHCP at (a) $f=1.55\text{GHz}$, (b) $f=3.8\text{GHz}$ and (c) $f=4.77\text{GHz}$.

Fig.5 shows the simulated farfield results at LHCP mode. At frequency 1.55GHz the main lobe magnitude is 6.3dBi, at frequency 3.8 GHz the main lobe magnitude is 3.7 dBi and main lobe direction is at 30.0 deg., at frequency 4.77 GHz the main lobe magnitude is 7.2 dBi and main lobe direction is at 5.0 deg.



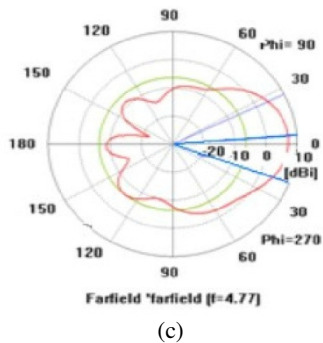


Fig.5 Simulated Farfield Pattern of LHCP at (a) $f=1.55\text{GHz}$, (b) $f=3.8\text{GHz}$ and (c) $f=4.77\text{GHz}$.

4. CONCLUSIONS

In this letter, a simple microstrip square patch with slotted ring inside the patch is proposed for dual band linear polarization and triple band circular polarization diversity. The circular polarization diversity is controlled by switching the current path in the patch using p-i-n diode configurations that is RHCP and LHCP. The experimental results show that the antenna is linearly polarized at 2.1GHz and 5.1 GHz and CP at three frequencies 1.54GHz, 3.8 GHz and 4.77 GHz. The multi band switchable antenna will find application in wireless communication in civilian as well as defence sector.

REFERENCES

- [1] N. Jin, F. Yang, and Y. Rahmat-Samii, "A novel patch antenna with switchable slot (PASS): Dual frequency operation with reversed circular polarizations," *IEEE Trans. Antennas Propag.*, vol. 54, no. 3, pp. 1031–1034, Mar. 2006.
- [2] A. Calmon, G. Pacheco, and M. Terada, "A novel reconfigurable UWB log-periodic antenna," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2006, pp. 213–216.
- [3] E. Erdil, K. Topalli, M. Unlu, O. A. Civi, and T. Akin, "Frequency tunable microstrip patch antenna using RF MEMS technology," *IEEE Trans. Antennas Propag.*, vol. 55, no. 4, pp. 1193–1196, Apr. 2007.
- [4] X.-S. Yang, B.-Z. Wang, and W. Wu, "Pattern reconfigurable patch antenna with two orthogonal quasi-Yagi arrays," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2005, vol. 2B, pp. 617–620.
- [5] Y.-F. Wu, C.-H. Wu, D.-Y. Sai, and F.-C. Chen, "A reconfigurable quadri-polarization diversity aperture-coupled patch antenna," *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 1009–1012, Mar. 2007.
- [6] S.-T. Fang, "A novel polarization diversity antenna for WLAN application," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2000, pp. 282–285.
- [7] M.-A. Kossel, R. Küng, H. Benedickter, and W. Bächtold, "An active tagging system using circular-polarization modulation," *IEEE Trans. Microw. Theory Tech.*, vol. 47, no. 12, pp. 2242–2248, Dec. 1999.
- [8] M. K. Fries, M. Gräni, and R. Vahldieck, "A reconfigurable slot antenna with switchable polarization," *IEEE Microw. Wireless Compon. Lett.*, vol. 13, no. 11, pp. 490–492, Nov. 2003.
- [9] A. Khaleghi and M. Kamyab, "Reconfigurable single port antenna with circular polarization diversity," *IEEE Trans. Antennas Propag.*, vol. 57, no. 2, pp. 555–559, Feb. 2009.
- [10] W.-L. Liu, T.-R. Chen, S.-H. Chen, and J.-S. Row, "Reconfigurable microstrip antenna with pattern and polarisation diversities," *Electron. Lett.*, vol. 43, no. 2, pp. 77–78, 2007.

A Reconfigurable Multiband Square Patch Antenna

Shweta Gautam¹, Ajit Yadav², Mithilesh Kumar³

^{1,2,3}Electronics Department,
University College of Engineering, Rajasthan Technical University, Kota
¹s.ecom.og@gmail.com, ²ajity2@gmail.com, ³mith_kr@yahoo.com

Abstract: The design of simple reconfigurable multiband square patch antenna is proposed, the antenna is excited by Microstrip feed line. The proposed antenna has square patch that is shorted to the ground plane through six shorting walls with six diodes. Three stubs with three diodes in each side of the square patch are present. The p-i-n diodes provide the opening and shorting of ground with the patch by taking the different combination of shorting diodes. Designed antenna radiate at a number of frequency bands, it shows single band, dual band and multi band operation. Key parameters have been analysed on Electromagnetic simulation software. The proposed antenna shows different dual band operations and multi band operation. Dual band occurs satisfactorily at 3.57GHz and 6.2GHz. Another dual band occurs satisfactory at 1.76GHz and 6.2GHz, and a multi bands occur at 3.06, 4.4 and 6.65 GHz.

Index Terms: Shorting wall, shorting p-i-n, multiband square patch, frequency reconfigurable antenna, dual band.

1. INTRODUCTION

In wireless communication patch antenna radiates for narrow bandwidth which limits its application [1]. So today's requirement is small size simple multiband or wide band antennas which are convenient to integrate in communication system. Wide band operation of antenna can be achieved by doing modification in structure of antenna. Some techniques are shorting wall [2], folded shorting wall [3], u-slot array [4], slots form [5], y-v slot [6], stacked patch [7], pair of slits on the patch (with total size of the antenna $150 \times 150 \text{ mm}^2$) [8], E-shaped patch on thick substrates with ground plane size of $140 \times 210 \text{ mm}$ [9] and using circular arc shaped slot on thick substrate [10].

This paper presents the frequency reconfigurable dual band and multi band operation of proposed antenna. This proposed antenna is multi tasking antenna with different frequency of operation. For achieving multi band operation patch structure is modified. The proposed antenna has square patch with microstrip feed line. This patch structure is modified by introducing the shorting wall in vertical sides of the patch which provide shorting between the patch and the ground. This shorting is achieved with help of p-i-n diodes, which provide different configuration and different shorting

ways between the patch and the ground. This antenna works satisfactory for single band, dual band and triple bands.

2. ANTENNA STRUCTURE

Fig.1 shows the geometry of the proposed reconfigurable multiband square patch antenna. Proposed antenna has three layer structure, bottom layer is a ground plane of thickness 0.1mm, on which there is a second layer of dielectric substrate FR4 of thickness 3 mm and dielectric constant is 4.9. The length of the ground plane of 70mm and the width is also 70mm. FR4 substrate too has 70 mm length and 70 mm width. The third layer is a square patch of length L.

The square patch with the shorting walls is shown in Fig1 (b) And fig1(c). There is six shorting walls 3 on each vertical sides of the square patch. These shorting walls are not directly shorting the ground plane to the patch; diodes are connected to provide desirable shorting between ground plane and the square patch. Fig.1 (b) shows the left view of the proposed antenna. There are three shorting walls in left side of the patch Shorting wall 1, shorting wall2 and shorting wall 3, and three p-i-n diodes D1, D2 and D3 are also shown in Fig.1 (b). Bottom of the shorting wall 1 is connected to the ground plane but top is not connected directly to the patch, top of shorting wall 1 is connected to the positive of the p-i-n diode D1 and negative of the p-i-n diode is connected to the patch.

In such a way top of the shorting wall 2 is connected to the patch and its bottom is not connected directly to the ground plane, there is a p-i-n diode D2 between the bottom of the shorting wall 2 and the ground plane. Similarly shorting wall 3 is also shown in Fig.1 (b). Top of the shorting wall 3 is connected to the patch via p-i-n diode D3.

The same structure is replicated in the right view of the proposed antenna shown in Fig.1(c) shorting wall 4 and shorting wall 6 are connected to the patch via diodes D4 and D6 respectively. And shorting wall 5 is connected to the ground plane via p-i-n diode D5.

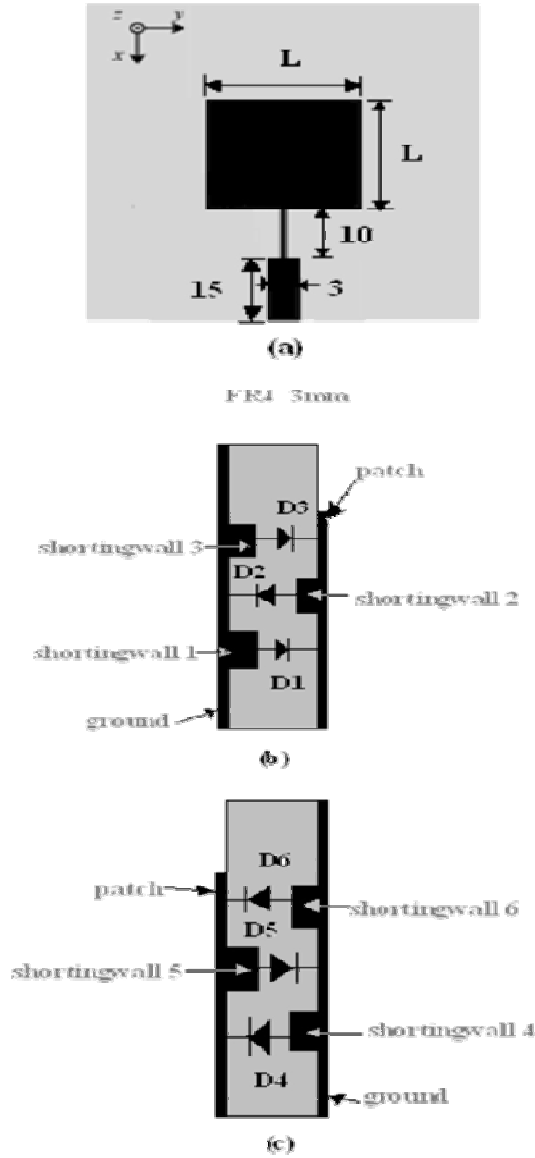


Fig.1. Geometry of reconfigurable and switchable dual band and multiband square patch antenna Dimensions: $70 \times 70 \text{ mm}^2$

(a) Top view of the antenna (b) Left Side view of antenna (c) Right Side view of antenna.

Square patch is feed by the Microstrip feed line. In Fig.1 (a) the $\lambda/4$ impedance transformer is used for impedance matching between the patch and the 50Ω feed line.

3. EFFECT OF DIODES IN ANTENNA PERFORMANCE

Table I shows the antenna performance for different biasing. When p-i-n diodes D1 and D3 are ON and D2, D4, D5 and D6 are OFF, in this case the shortingwall-1 and shortingwall-3 shorts the ground plane and the patch, other

shorting walls are open. In this case the proposed antenna radiates at two frequencies i.e. 3.57 and 6.2 GHz, Antenna worked as dual band antenna. Other case when diodes D2 and D5 are ON and all other diodes (D1, D3, D4, and D6) are OFF shortingwall-2 and shortingwall-5 shorts the ground plane and the patch and other shorting walls are open. In this case the proposed antenna radiates at two different frequencies 1.76 and 3.32 GHz and still work as a dual band antenna.

Table-I: Effect of Diode Switching

		Diode state				Radiatings	
D1	D2	D3	D4	D5	D6	F1	F2
ON	OFF	ON	OFF	OFF	OFF	F1 = 3.57 GHz	F2 = 6.2 GHz
OFF	ON	OFF	OFF	ON	OFF	F1 = 1.76 GHz	F2 = 3.32 GHz
ON	OFF	ON	ON	OFF	ON	F1 = 2.68 GHz	F2 = 4.32 GHz
OFF	OFF	OFF	OFF	OFF	OFF	F1 = 3.13 GHz	F2 = 6.65 GHz
ON	ON	ON	ON	ON	ON	F1 = 3.06 GHz	F2 = 4.4 GHz
						F3 = 6.2 GHz	

But when diodes D1,D3,D4,D6 are ON and other (d2 and d5 are OFF), shortingwall-1,shortingwall-3,shortingwall-4 and shortingwall-6 shorts the ground plane with the patch and in this case the proposed antenna radiate at three frequency bands 2.68,4.32 and 6.2GHz and antenna work as a triple band antenna. Proposed antenna also work as single band antenna if all diodes are in OFF state in this case there is no shorting between ground plane and patch and proposed antenna radiate at 3.13 GHz. But when all the diodes are in OFF state this antenna again shows response as a triple band antenna.

4. IPLIMENTATION AND RESULTS

Proposed antenna has square patch of length $L=20\text{mm}$, FR4 substrate of $\epsilon_r=4.9$, and thickness of 3mm, lenght of 50Ω microstrip line is 10mm and $\lambda/4$ quarter wave transformer length is 15mm to match the feed line to the patch. Shorting walls length is 2mm and width is 1.5mm. The patch and the ground have thickness of .01mm.

The proposed antenna work as single band, dual band and multi band antenna by shorting the differed shorting walls via p-i-n diodes .A few cases are studied and shown in this paper.

Case-1, when diodes D1, D5 are ON and other diodes are OFF in this case antenna work as dual band antenna and radiate at frequencies 3.57 and 6.2 GHz. Fig.2 shows the s-parameter verses frequency curve of Case-1. Case-2, when diode D2 and D5 are ON and other diodes are OFF, proposed antenna still work as dual band antenna but two frequencies are different from case-1 that is 1.76 and 3.32 GHz.

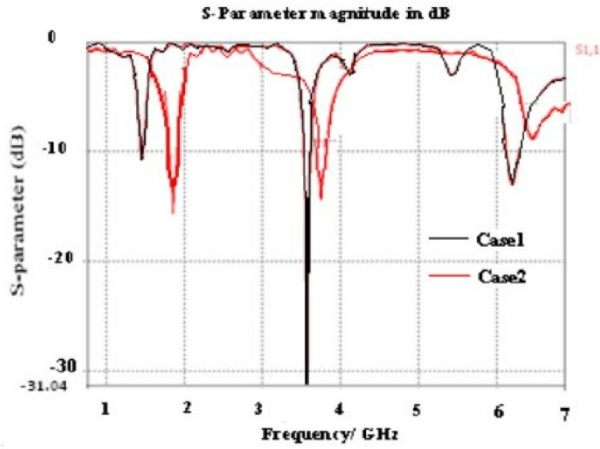


Fig. 2 Simulated s-parameter of the proposed antenna when

Case-1(D1=ON, D5=ON) Case-2(D2=ON, D5=ON)

Case-3, when D1, D3, D4, D6 are in ON state and D2 and D5 are OFF antenna shows the triple band operation shown in Fig.3. Three bands are at 2.68 GHz, 4.32 GHz and 6.2 GHz.

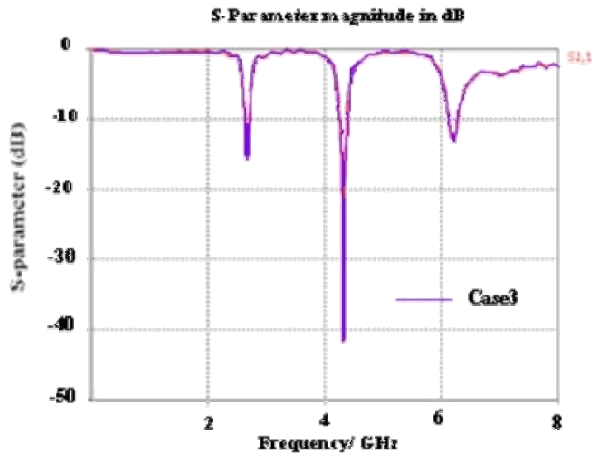


Fig. 3. Simulated s-parameter of the proposed antenna when Case-3 (D1=ON, D3=ON, D4=ON, D6=ON), Triple band.

Case-4, when all diodes are in OFF state, there is no shorting between ground plane and patch. In this case antenna work as single band antenna shown in Fig.4 and radiate at frequency 3.13GHz.

Case-5, when all diodes are ON again antenna radiate triple band shown in 3.06GHz, 4.4 GHz, 6.65 GHz. Some other cases are also possible.

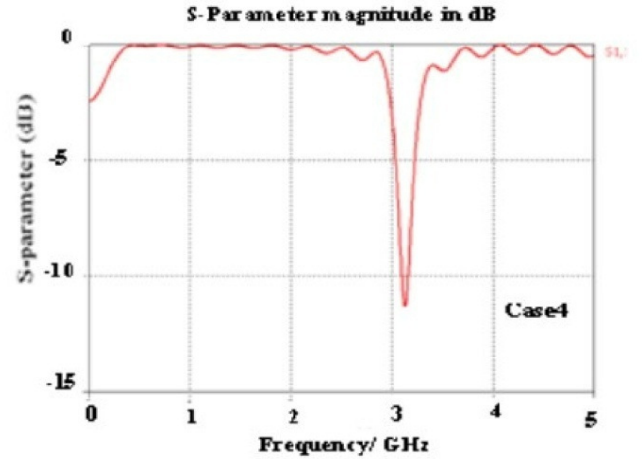


Fig. 4. simulated s-parameter of the proposed antenna when Case-4 (All diodes are OFF, single band operation)

Table II: Multi Band Operation y A Single Antenna

CASES	SHORTING/OPENING	MULTIBAND
	OFF DIODES	
Case-1	D1,D5 ON	Dual band antenna
	D2,D4,D5,D6 OFF	
Case-2	D2,D5 ON	Dual band antenna
	D1,D3,D4,D6 OFF	
Case-3	D1,D3,D4,D6 ON	Triple band antenna
	D2,D5 OFF	
Case-4	All diodes are OFF	Single band antenna
Case-5	All diodes are ON	Triple band antenna

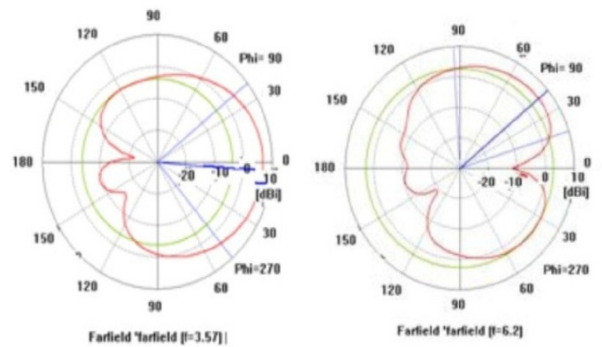


Fig. 5. Simulated Far field Pattern of case-1, at 3.57GHz and 6.2GHz.

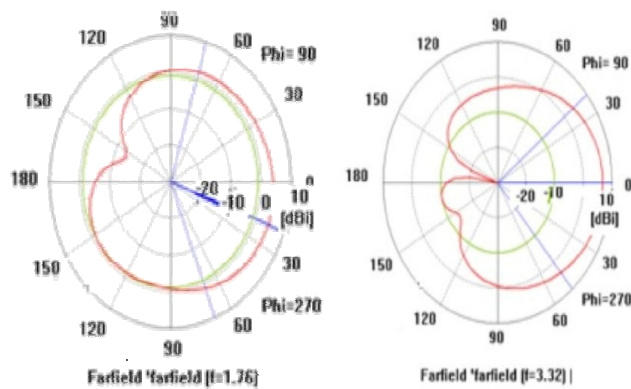


Fig. 6. Simulated Far field Pattern of case-2, at 1.76GHz and 3.32GHz.

The Fig.5 shows the far field pattern of case-1 shown in Table-II. Far field pattern are studied at 3.57GHz and 6.2GHz, at frequency 3.57GHz the main lobe magnitude is 6.8dBi and main lobe direction is 5.0 degree. At frequency 6.2GHz the main lobe magnitude is 5.8 dBi and main lobe direction is 40.0 degree. Fig .6 shows the far field pattern of case-2 shown in Table-II. Due to dual band operation far field pattern are studied at 1.76GHz and 3.32 GHz. At 1.76GHz the main lobe magnitude is 4.4dBi and the main lobe direction is at 20.0 degree .and at 3.32GHz the main lobe magnitude is 6.8dBi and the main lobe direction is 0.0 degree.

5. CONCLUSIONS

In this letter, a simple reconfigurable multiband square patch antenna is proposed. The dual band and multi band operation of antenna is taken by shorting the ground plane and the patch through shorting walls and p-i-n diodes. By making the diodes ON and OFF different configuration are taken and due to these different configuration dual band and tripal bands are observed. So by doing proper biasing the diodes,

dual band reconfigurable multiband are achieved. The experimental results shows that one dual band occur at frequency 3.57 GHz and 6.2 GHz. Second dual band occur at 1.76 GHz and 6.2 GHz. And the triple band occurs at 3.06, 4.4 and 6.62 GHz successfully by electromagnetic simulation software.

REFERENCES

- [1] G. Kumar and K. P. Ray, *Broadband Microstrip Antennas*. Boston, MA: Artech House, 2003, pp. 18–23.
- [2] C. Mak, R. Chair, K. Lee, K. Luk, and A. Kishk, “Half U-slot patch antenna with shorting wall,” *Elect. Lett.*, vol. 39, pp. 1779–1780, 2003.
- [3] Y. Li, R. Chair, K.M. Luk, and K. F. Lee, “Broadband triangular patch antenna with a folded shorting wall,” *IEEE Antennas Wireless Propag. Lett.*, vol. 3, no. 1, pp. 189–192, Dec. 2004.
- [4] H. Wang, X. B. Huang, and D. G. Fang, “A single layer wideband U-slot microstrip patch antenna array,” *IEEE Antennas Wireless Propag. Lett.*, vol. 7, pp. 9–12, 2008.
- [5] Y. Lee and J. Sun, “A new printed antenna for multiband wireless applications,” *IEEE Antennas Wireless Propag. Lett.*, vol. 8, pp. 402–405, 2009.
- [6] S. Qu and Q. Xue, “A Y-shaped stub proximity coupled V-slot Microstrip patch antenna,” *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 40–42, 2007.
- [7] J. Anguera, C. Puente, C. Borja, and J. Soler, “Dual-frequency broadband- stacked microstrip antenna using a reactive loading and a fractalshaped radiating edge,” *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 309–312, 2007.
- [8] K.-L. Wong and W.-H. Hsu, “A broad-band rectangular patch antenna with a pair of wide slits,” *IEEE Trans. Antennas Propag.*, vol. 49, no. 9, pp. 1345–1347, Sep. 2001.
- [9] F. Yang, X. Zhang, X. Ye, and Y. Rahmat-Samii, “Wide-band E-shaped patch antennas for wireless communications,” *IEEE Trans. Antennas Propag.*, vol. 49, no. 7, pp. 1094–1100, Jul. 2001.
- [10] R. Bhalla and L. Shafai, “Broadband patch antenna with a circular arc shaped slot,” in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2002, vol. 1, pp. 394–39.

Design of Compact BPF for UWB Communication using Multi-Physics

Malabi Singh¹, Mihir Narayan Mohanty²

^{1,2}*School of Electronics Engineering,
ITER, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India*
¹*malabi.singh@gmail.com,* ²*mihir.n.mohanty@gmail.com*

Abstract: There is an increasing demand in favour of RF, microwave, and millimetre wave filters with more stringent necessities. Compact and broadband bandpass filter is a passive constituent and highly demanded in a UWB technique. On microstrip building, can provide the advantages of simple design, low cost, compact size, and is widely used. This paper proposes a design of bandpass filter to built on microstrip using GaAs substrate. The design originates from modelling the series stub as a technique of two disproportionately coupled transmission lines, as it is equivalent to a basic filter element of admittance inverter. The presented ideal is very compact in relation to its resonant frequency and provides a relatively eminent Q-factor compared to capacitively coupled microstrip line ideal designs and besides is comparable with Si substrate device.

Index Terms: Coplanar Waveguide (CPW), Bandpass Filter (BPF), ultra wideband (UWB), GaAs, Microstrip.

1. INTRODUCTION

Electronic filters are circuits so as to hold indicatives giving out functions. They transform an input indicative to attain an output indicative with the obligatory characteristics. In vogue the frequency domain filters are used to rebuff not needed indicate frequencies and to pass signals of desired frequencies. Filters are indispensable procedure, in many systems and applications as well as wireless broadband, mobile, and satellite communications, radar, navigation, sensing and other systems. With the development of these systems, mostly induced by grand business-related interests, narrow electromagnetic spectrum has to be shared amid more and more systems. Thus, nearby is an increasing demand in favour of RF, microwave, and millimetre wave filters with more stringent chuck filters are employed in various systems to cliquey or confine signals surrounded by specified spectral limits.

COMSOL Multiphysics is a software environment for the modelling and simulation of any physics-based system. A particular strength is its ability to account for multiphysics phenomena. Optional modules add discipline-specific tools for mechanical, fluid, electromagnetics, and chemical simulations, as well as CAD interoperability.

The paper is organized as follows. Section 1 introduces the work. Section 2 describes the related literature. Section 3 provides a general procedure of design with COMSOL. Section 4 explains the section 3 with suitable BPF. Section 5 is for result of the work and finally section 6 concludes the work.

2. RELATED LITERATURE

As the system operating frequency becomes higher and higher, CPW's are increasingly used as transmission lines in MIC and MMIC design. Literature suggests, some of the works have been done on BPF [1-4], and HTS BPF based cryogenic receiver front-ends have been designed as well as the CPW structure is more advantageous than the micro-stripline structure because of only one side HTS coating and easy for size reductions. In some studies, also authors designed the miniaturized cross-coupled CPW BPF by using highly packed meander line half-wavelength ($\lambda/2$) resonators and inter-digital space.

Owing to the uni-planar feature, CPW offers several advantages over its counterpart microstrip line, such as easy series and shunt connection, no via hole, insensitive to the substrate thickness, and low dispersive effects. In spite of these advantages, applications of CPW, especially for filters, are not as widely as expected because the lack of design data and accurate equivalent circuit model is difficult to design. In 1976, Houdart [1] proposed the CPW open-end and short-end series stubs and which may be used in the filter design.

The two stubs were modelled as a series capacitor and inductor respectively, which are too simple to predict properly the electrical properties. Dib *et al.* applied the space-domain full-wall wave integral equation to calculate the scattering parameters of the stubs [2]. From the computation results, they used curve-fitting technique to get the equivalent circuit models. Their models can provide better accuracy but the elements in the models are too complicated to convey reasonable physical meaning. Similarly, Ray *et al.* [3] used the simulator *emto* to calculate the scattering parameters but suggested another equivalent circuit models. Since the circuit model is not enough to realize the

filter, they also employed the open-end or short-end shunt stubs. As a result, filter consumes a large layout area and requires additional air bridges.

A few structures were experimentally employed and presented to implement the CPW filter. In 1983, Walliam *et al.* [4] proposed a CPW end-coupled filter using CPW gap interdigital capacitances as admittance inverters to realize the filter. However, the gap capacitance is usually not large enough for the design of filters, especially wideband bandpass filters. Nguyen [5] suggested the broadside-coupled coplanar waveguide to achieve wide-bandwidth but the main advantages due to the uni-planar feature of CPW will be destroyed. Recently, Chang *et al.* [6] used the CPW shunt inductor as the impedance inverter to realize the band pass filter. The filter occupies larger area due to the shunt stubs and consequently, is not satisfactory in MIC and MMIC which always require high density to reduce the circuit cost.

The Federal Communications Commission (FCC) released a frequency band 3.1-10.6 GHz as ultra-wideband (UWB) commercial communications and that has been used as a standard. Attention has been paid to applications of ultra-wideband (UWB) technology on wireless communication system. In this technology local area networks, position location and tracking, and radar systems can be considered as some of the applications. Many UWB devices with high data transmission rate and very low power consumption circuits have been proposed and investigated widely. It is a challenge to reduce their size and weight in order to integrate them with other components as a compact system. Microstrip filters are one of the most popular realizations of planar microwave filters [7-12].

Now-a-days, many novel microstrip and other planar filters with advanced filtering characteristics are developed using novel materials and fabrication technologies such as HTS, liquid crystal polymers (LCP), LTCC, MMIC, and microelectromechanical systems (MEMS). These filters as well as advanced filters built using conventional Alumina or Duroid substrates are designed using novel CAD tools.

Gallium Arsenide is superior to those of silicon. It has a higher saturated electron velocity and higher electron mobility, allowing gallium arsenide transistors to function at frequencies of 250 GHz. Unlike silicon junctions, GaAs devices are relatively insensitive to heat owing to their wider bandgap. Also, GaAs devices tend to have less noise than silicon devices, especially at high frequencies. These advantages recommend GaAs circuitry in mobile phones, satellite communications, microwave point-to-point links and higher frequency radar systems and is verified in this work.

3. MODEL DESIGN

Microstrip is the most popular planar transmission structure used in MIC. Planar transmission structure is the one in which the characteristics of the circuit elements, built using this structure, can be determined by the dimensions in a single plane. This is the main requirement for a transmission line to be used in MIC. Microstrip can be fabricated using photolithographic processes. Open configuration makes it easily integrated with other discrete lumped passive and active microwave devices. Microstrip transmission lines consist of a conductor printed on top of thin, grounded dielectric substrate. The width of the conductor w , thickness of the substrate h , and relative permittivity ϵ_r are the main important parameters. Thickness of the top metallic line t is much less important and often can be neglected. It can be easily customized the layout of COMSOL [13].

Bandpass filters are used primarily in wireless transmitters and receivers. The main function of such a filter in a transmitter is to limit the bandwidth of the output signal to the minimum necessary to convey data at the desired speed and in the desired form. In a receiver, a bandpass filter allows signals within a selected range of frequencies to be heard or decoded, while preventing signals at unwanted frequencies from getting through. A bandpass filter also optimizes the signal-to-noise ratio (sensitivity) of a receiver.

In both transmitting and receiving applications, well-designed bandpass filters, having the optimum bandwidth for the mode and speed of communication being used, maximize the number of signals that can be transferred in a system, while minimizing the interference or competition among signals. The basic circuit diagram of BPF is shown in Fig.1.

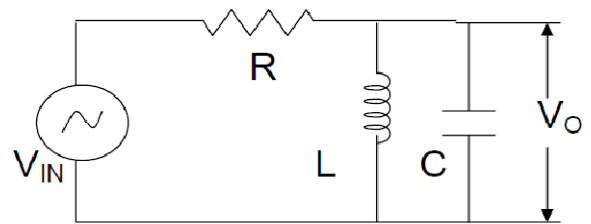


Figure 1: Circuit Diagram for a Band Pass Filter.

The technological process involved in this paper deals with a CPW micro-machined GaAs based technology derived from the work. Coplanar waveguide (CPW) bandpass filters can be realized using inter-digital capacitors (IDCs) and short-circuited stub inductors (SSIs). Such a filter can readily be implemented on a GaAs wafer. The presented model is very compact ($0.07\lambda \times 0.05\lambda$) in relation to its resonant frequency and provides a relatively high Q-factor compared to capacitively coupled microstrip line model designs.

The structure can be realized by etching a pattern in a thin gold layer on a high dielectric ($\epsilon_r = 12$) GaAs substrate. In this model, the gold layer is treated as an infinitely thin layer of perfectly conducting material. Two lumped ports represent a coplanar waveguide coupling into, and out of, the device. The lumped port applies a voltage difference between the center conductor and the ground planes. This voltage difference is applied by adding a small metallic air bridge that equally divides the signal between the two ground planes. The line width and coupling gap on the comb in the interdigital capacitor is 50 microns, which is wide enough to account for etching tolerances. Series and parasitic SSIs are added to generate a bandpass frequency response.

The model space consists of the GaAs wafer, with the pattern on the surface, and an air box surrounding the entire structure. The air box, in turn, is bounded by a perfect electric conductor boundary representing the packaging placed far enough from the CPW so as not to introduce any unwanted coupling.

4. RESULT

Here the model is designed for a coplanar waveguide (CPW) bandpass filter which is realized using inter-digital capacitors (IDCs) and short-circuited stub inductors (SSIs). It is very compact (0.07λ -by- 0.05λ) in relation to its resonant frequency and provides a relatively high Q compared to capacitively coupled microstrip line model designs. The structure is simulated over a range of frequencies from 3.7 GHz to 3.9 GHz. The simulation results show bandpass filter characteristics around 3.8 GHz as presented in Figure 3.

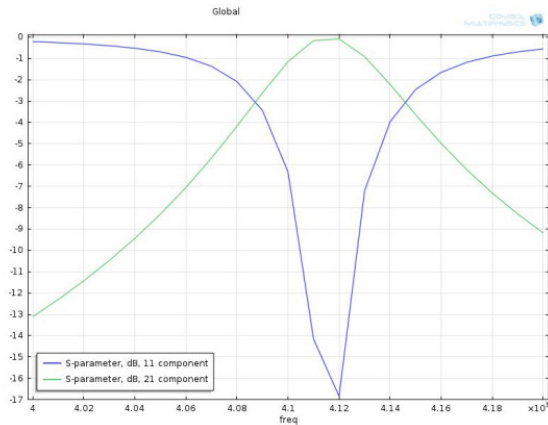


Fig. 3. The frequency response of the coplanar waveguide filter shows bandpass characteristics using Silicon

The Q -factor, defined as center frequency / (-3 dB bandwidth), evaluates to approximately 78. Because the filter is enclosed by the PEC package, it is effectively a

grounded coplanar waveguide circuit and there is a parasitic reactance between the circuit and the bottom ground plane. This reactance loading lowers the Q -factor, which is generated from the circuit itself. The model without the package as well as the bottom ground plane provides sharper frequency responses and a Q -factor above 150.

Though air bridges are used only for the port excitation in this model, it is generally recommended to add air bridges around SSIs when structures of this type are used for very high frequencies in order to suppress any potential radiation over the slots.

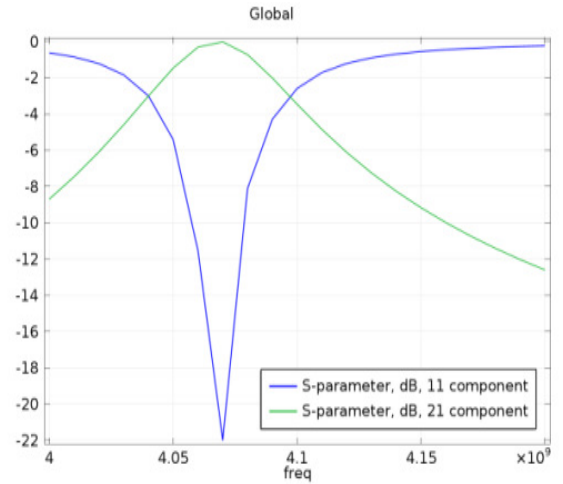


Fig. 2. The frequency response of the coplanar waveguide filter shows bandpass characteristics using GaAs.

GaAs has an energy gap that is four orders of magnitude larger than Si. This allows GaAs to be made semi-insulating (with a bulk resistivity on the order of 10^9 ohms). Devices made in semi-insulating GaAs substrates have reduced parasitic capacitance which leads to further improvement in speed over silicon.

GaAs has the ability to emit light which is useful for making lasers, light-emitting diodes, and microwave emitters used in cellular phones.

Table 1: (Using Si Wafer)

Name	Expression	Description
W_center	200[um]	CPW center line width
W_gap	125[um]	CPW gap
H_wafer	200[um]	Wafer thickness
f_min	3.4[GHz]	Minimum frequency in sweep
f_max	3.8[GHz]	Maximum frequency in sweep
h_max	$0.2 * c_{const} / f_{max}$	Maximum mesh element size, air

Table-2(UsingGaAs Wafer)

Name	Expression	Description
W_center	150[um]	CPW center line width
W_gap	125[um]	CPW gap
H_wafer	150[um]	Wafer thickness
f_min	4.2[GHz]	Minimum frequency in sweep
f_max	4.6[GHz]	Maximum frequency in sweep
h_max	$0.2 * c_const / f_max$	Maximum mesh element size, air

5. CONCLUSION

The microstrip filters can be used on wide range of frequency bands by employing various kinds of substrate materials. The main disadvantage of this type of filter is high insertion loss, due to significantly lower Q factor. The rapid development of microstrip and other planar filters is to be taken care of new materials and fabrications technologies. Using this COMSOL environment, the design based on MEMS can be performed. The comparison among Si and GaAs based design is shown in Fig 2 and Fig 3. The conclusion can be drawn as both discrete components and integrated circuits made in GaAs has higher frequency than those made of silicon because its low-field electron mobility is larger than that of silicon, and GaAs has a lower saturation field than silicon.

REFERENCES

- [1] M. Houdart, "Coplanar lines: Application to broadband microwave integrated circuits," in *Proc. 6th Euro. Microwave Conf*, Rome, 1976, pp. 49-53.
- [2] N. I. Dib, L. P. B. Katehi, G. E. Ponchak, and R. N. Simons, "Theoretical and experimental characterization of coplanar waveguide discontinuities for filter applications," *IEEE Trans. Microwave Theory Tech.*, vol. 39, pp. 873-881, May 1991.
- [3] A. K. Rayit and N. J. McEwan, "Coplanar waveguide filters," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1993, pp. 1317-1320.
- [4] D. F. Williams and S. E. Schwarz, "Design and performance of coplanar waveguide band-pass filters," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-31, pp. 558-566, July 1983.
- [5] C. Nguyen, "Broadside-coupled coplanar waveguide and their endcoupled band-pass filter applications," *IEEE Trans. Microwave Theory Tech.*, vol. 40, pp. 2181-2189, Dec. 1992.
- [6] C. Y. Chang, H. K. Chiou, T. H. Wang, and C. C. Chang "A CPW inductor coupled bandpass filter," in *Asia Pacific Microwave Conf*, 1993, vol. 2, pp. 16.69-16.73.
- [7] R. J. Cameron, C. M. Kudsia, and R. R. Mansour, *Microwave filters for communication systems: fundamentals, design, and applications*. Hoboken, New Jersey: John Wiley & Sons, 2007.
- [8] M. N. S. Swamy and K.-L. Du, *Wireless Communication Systems: From RF Subsystems to 4G Enabling Technologies*, New York: Cambridge University Press, 2010.
- [9] H. Howe, "Microwave Integrated Circuits: An Historical Perspective," *IEEE Trans. on Microwave Theory and Tech.*, vol. 32, no. 9, pp. 997- 1008, September 1984.
- [10] J. G. Hong and M. J. Lancaster, *Microstrip Filters for Rf/Microwave Applications*, New York: John Wiley & Sons, 2001.
- [11] R.E. Collin, *Foundations for Microwave Engineering*, McGraw-Hill, 1992.
- [12] S. Mao and others, "Modeling of Symmetric Composite Right/Left-Handed Coplanar Waveguides with Applications to Compact Bandpass Filters," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 11, Nov. 2005.
- [13] [www.comsol.co.in/comsolmultiphysics4.3/model library](http://www.comsol.co.in/comsolmultiphysics4.3/model%20library)

Target Position Estimation by Synthetic Aperture Radar (SAR) Dataset

Ganesh Dutt¹, Kolli Sridatta Sairam Reddy², Ishan Sharma³,
Sudhir Kumar Chaturvedi⁴, Ugur Guven⁵, Pavan Kumar Nanduri⁶

^{1,2,3}*B. Tech Avionics Engineering Students Department of Aerospace Engineering
University of Petroleum & Energy Studies, Dehradun, India*

¹*mail2ganeshdutt@gmail.com*

^{4,5,6}*Faculties, Department of Aerospace Engineering
University of Petroleum & Energy Studies, Dehradun, India
sudhir.chaturvedi@ddn.upes.ac.in*

Abstract: SAR is a well-established remote sensing technique that can provide high resolution radar images from space independent of weather and sunlight illumination. This paper presents the preliminary technique to retrieve the target positions using space-borne SAR and ship-borne Automatic Identification System (AIS). In the SAR images with high resolution, each target occupies several resolution units to form area target. Detecting the vessel target in the SAR images with high resolution should regard the target as distributive targets.

The main objective of this work consists of two main stages: The first is to detect the targets near to the port using the SAR and the second is to identify and match the detected targets using AIS with their respective position.

The proposed technique mainly deals with the main stages such as acquisition of SAR image, sub-image extraction from the image, image threshold method, dilation and erosion processes, concatenating operation over detected targets. The AIS datasets will be used for the matching of the detected target positions to obtain the good estimates of the result. AIS data is the series of the datasets which consists of various information such as vessel name, Maritime Mobile and Service Identification (MMSI) numbers, coarse over ground, speed over ground, expected date of arrival and departure, latitude and longitude positions.

The study will be carried out based on the TerraSAR-X image data acquired over the Tokyo Bay on May 30, 2011. This study result will provide an important contribution for designing near-real-time operational system for the monitoring of the vessel targets near to coast as well in open sea waters. Also the proposed method may be useful for the detecting the vessel targets in coastal or deep sea to resolve maritime safety and security problems over the huge oceanic regions. SAR-AIS correlated vessels can be declared as the friend target and no correlated targets can be assigned as the foe targets.

Index terms: SAR, AIS, Target, Image processing, Dilation

1. INTRODUCTION

The vital role of air borne radars to extract information and detection of targets came into prominent usage in the fields of environmental monitoring, earth-resource mapping, and military systems from the past 4 decades. Since then various radar systems have been implemented to extract information and track while scan (detect) the target(s) position. SLAR was firstly used for detection of the target position which has a spatial resolution depending on the altitude of a platform. For the better spatial resolution SLAR was substituted by Synthetic Aperture Radar (SAR) also known as Imaging Radar (1).

The objective of this paper is to detect the accurate coordinates of the target using SAR data which is extracted from the region of Tokyo bay in Japan. The lower Earth Orbit satellite, TerraSAR-X which operates in X band frequency range under German Aerospace Centre (DLR) has been used to extract this data set. The primary problem faced is the noise which degrades the quality of the SAR image known as SPECKLE. Usually it increases the mean Grey level of the Region of Interest (ROI) causing difficulties in the image interpretation as explained in the paper (2). The vessel target detection by means of integration of SAR and AIS has also been carried out using various matching method based on the vessel signal signature parameter, position, velocity and size matching techniques (3).

The raw SAR data captured by TerraSAR-X (TSX) for the time duration of 2 seconds i.e. GMT 08:36:33-08:36:35 and local time in Japan, Tokyo bay i.e. 17:36:33-17:36:35 on 30th day of May, 2011. The scan centred coordinates are lat 35.2800° long 139.7720° and c_{centered} is 38.723° and HH polarized which is having a total image size of (6415x7217) bytes.

2. METHODOLOGY

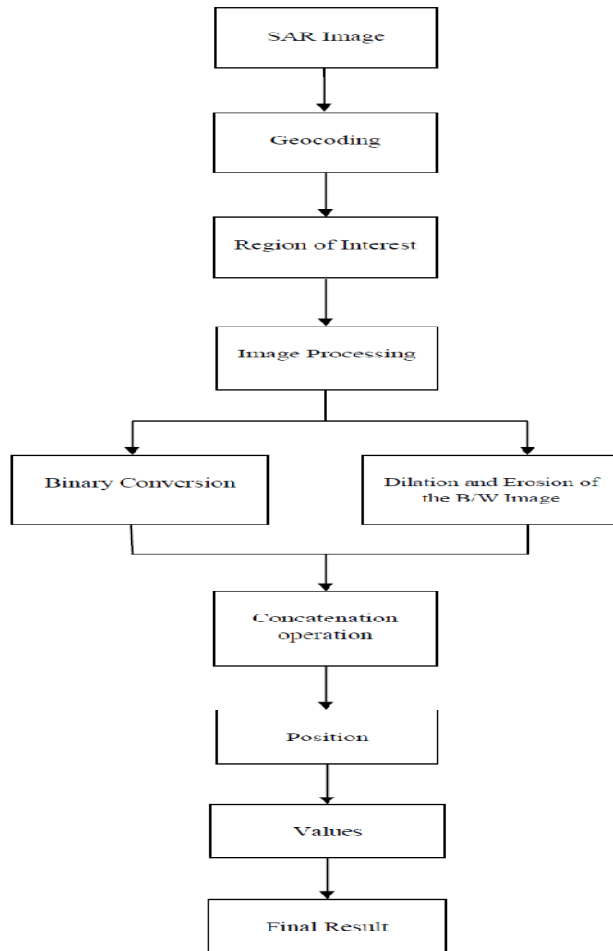


Fig. 1. SAR target detection methodology

Figure 1 shows the basic block diagram of the target detection methodology. SAR images were geocoded with respect to the earth's coordinate. A small image region of interest was selected for the processing of an image. The detailed about the SAR specifications are given in Table 1. SAR image always gives the data information in terms of greyscale image. For detection of the brighter pixels in image, image thresholding method was applied which converts the greyscale (0-255) to binary image (0-1).

The output window result shows the various noise presences in image due to high backscattering response from the ocean surface. Dilation operation was then performed which works on the principle that, it combines all the small pixels into group and gives the output result with certain signal signature boundaries (4,5). The erosion process was performed to remove the unwanted clutter and noises present in image. In order to estimate the positions, concatenating matrix operation can be applied which find out the centroid of each pixels and the centre positions can be converted from image

pixels to geographic latitude and longitude positions using equations 1 and 2.

Table I. SAR Input Parameters and its respective data information

Parameters	Attributes
Satellite	TerraSARX
Lower Earth Orbit	750 Km
Velocity	7.07 Km/sec
Polarization	HH
Location	Japan, Tokyo Bay
Date	30 th May, 2011
Time(GMT)	08:36:33-08:36:35
Duration	2 seconds
Latitude	35.2800°
Longitude	139.7720°
$\theta_{centered}$	38.723°
Data size	6415 x 7217 bytes
Operating Frequency	X band

A. Equation for Position Estimation

The positions of the target can be estimated using the equations (1) and (2).

$$X_{max} = X_{min} + (N_x - 1) * \Delta X \quad (1)$$

$$Y_{max} = Y_{min} + (N_y - 1) * \Delta Y \quad (2)$$

Where, X_{max} = maximum longitude of the ROI image, X_{min} = Longitude minimum from Geo referencing matrix, N_x = range of lines, ΔX = Geo referencing matrix along Longitude, Y_{max} = Maximum latitude of the ROI image, Y_{min} = Minimum latitude of the ROI image, N_y = range of samples, ΔY = Geo referencing matrix along sample axis

3. RESULTS

The final output screen is as shown in Fig (4) which came out of the dilation Process and also it represents the 4 detected brighter signal targets at various positions.

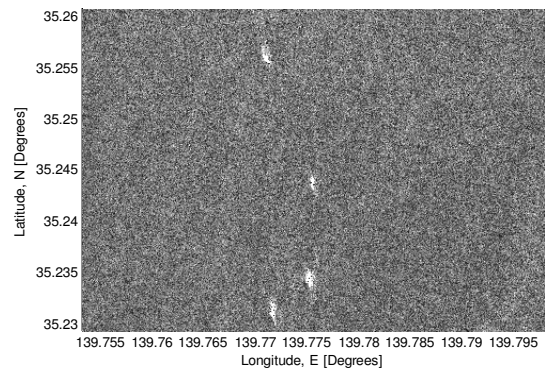


Fig. 2 Detection of targets at Tokyo Bay, Japan

Figure 2 represents sub-image area of Tokyo Bay, Japan which consists of various ship targets with various clutters present in the image.

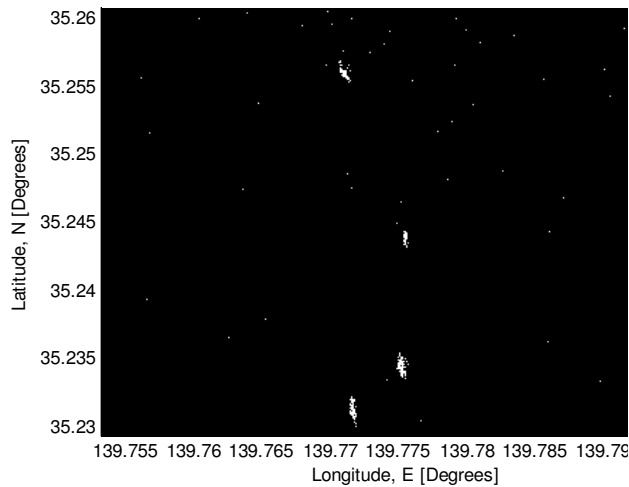


Fig. 3. Grayscale image converted to Binary image with noise.

Figure 3 represents Image Processing result with the presence of noise and detected targets.

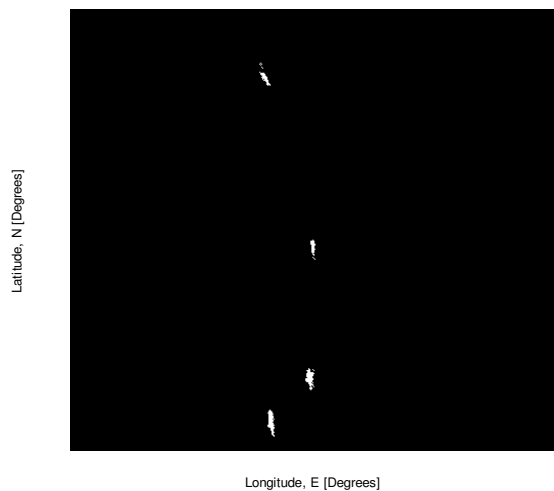


Fig. 4. Output screen after dilation and erosion process, it shows the 4 detected targets at various positions.

Figure 4 is the final output screen for detected ships targets with detailed signatures over the screen. The background is completely dark represents the low backscattering response while the brighter target give response of high signal signature sensed by SAR.

4. CONCLUSION

SAR data has a prominent role in detection and positioning of target (s). The technique used in this paper to retrieve the vessel target position and reduction of noise produced in sea clutter conditions is achieved by using Envi and MatLab softwares. The result shows the different types of vessel targets.

5. ACKNOWLEDGEMENT

The author (s) would like to thank University of Petroleum & Energy Studies, Dehradun for providing the opportunity to write this research article.

The Corresponding author also would like to thank Dr. C.S. Yang, Korea Ocean Research & Development Institute, Korea (Republic of) for providing the SAR dataset over the coast of Tokyo Bay, Japan.

REFERENCES

- [1] G.Russo, G.Ferrara, M.Migliaccio, A.Montuori, F.Nunziata and A.Sorrentino, "A K-Generalized Speckle Model based Indicator for Ship Detection over SLC SAR Images", Università degli Studi di Napoli Parthenope, Centro Direzionale, Isola C4, 80143 Napoli, Italy.
- [2] S.K. Chaturvedi, C.S. Yang, K. Ouchi, and P. Shanmugam "Ship Recognition by SAR and AIS," International Journal of Navigation. vol. 65, pp. 323–337, March 2012.
- [3] C. R. Jackson, and J. Apel "Synthetic Aperture Radar Marine User's Manual", Chapter 1. Principles of Synthetic Aperture Radar, 1-24 NOAA, Washington, 2004.
- [4] B. Camilla, D. J. Weydahl, R. Olsen, "Ship traffic monitoring using multi-polarization satellite SAR images combined with AIS reports", Norwegian Defence Research Establishment (FFI), 2008.
- [5] O. Lavrova, S. Shcherbak, M. Mityagina, V. Pyrkov, "Case Study: Use of SAR Data for the Operational Control of Fishing Ships Positioning", Envisat Symposium 2007, Switzerland, 23-27 April 2007.

Anthropogenic Treaty

Shefali Nagar¹, Shikha Puri², Priyanshi Dwivedi³

^{1,2,3}Department of Electronics and Communication Engineering,
Gautam Buddh Technical University (GBTU)

Raj Kumar Goel Institute of Technology for Women
Ghaziabad, Uttar Pradesh – 201003, India

¹shefalinagar@gmail.com, ²shikhapuri982@gmail.com

Abstract: All types of Telecommunication depends on satellite. A communication is a process of establishing connection or link between to points for information exchange throughout everywhere. The process of sending message like radio telephone, Radio broadcasting, TV Broadcasting, TELAX and Fax with the help of satellite station in our space. Communication is mainly done through geostationary satellite. The communication satellite has some devices called “transponders” and transmits them again in different direction. The audio or video (pictures) signal transmitted by an earth station like a TV station in the form of EM waves. In all India transmission of television programme of Doordarshan has been made possible through INSAT satellite. To keep a continuous track of various types of communication like long distance telephone calls, TELAX (Printed message) and FAX (Picture or Printed matter) for better result of Telecommunication. We have to maintain all environment condition as global warming

1. INTRODUCTION

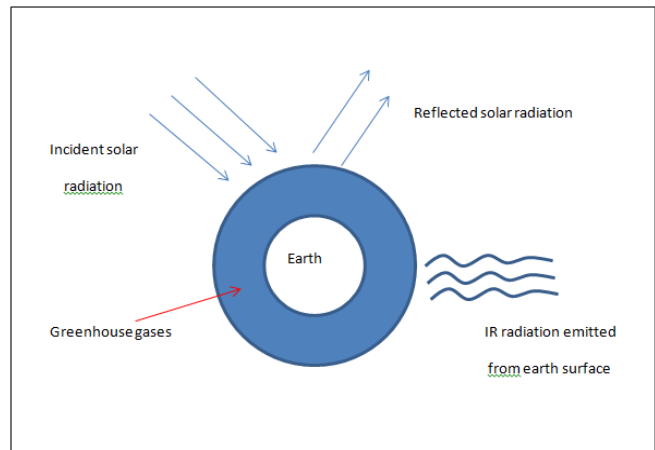
We are always searching effective ways to improve the environmental conditions. In the context of environmental policy, climate change usually refers to changes in modern times it may be qualified as anthropogenic (by population) climate change more generally known as “global warming” or anthropogenic global warming (AGW). Today the most crucial environmental related issue is climate change or global warming. In the recent past, global warming observation home provided a clear evidence of climate changes resulting from anthropogenic activities. The Earth’s temperature has risen by 0.5°C over the past century and recent years have been among the hottest on record. Human activity has influence the climate in at least three ways.

1. By changing the radiation at properties of the Earth’s surface
2. By venting waste heat into atmosphere.
3. By changing the concentration of key gaseous components of the atmosphere. Trace gases or greenhouse gases like CO₂, ClO_x, CH₄, N₂O, HO_x, O₃, CFC_s etc are minor gaseous constituents of the

atmosphere. However they play a surprisingly dominate role in regulating the entire earth atmosphere.

2. GREEN HOUSE EFFECT

The sun heats the earth. Solar radiation pass through the atmosphere and is most absorbed at earth’s surface. Only small portion is reflected back into space. The absorbed heat is emitted from the surface as infra-red radiation. Fortunately this radiated heat (in the form of IR radiation) cannot escape from the atmosphere of the earth. Same of this radiated energy is “trapped” by a number of gases present in the atmosphere. This phenomenon of “heat allowed in but cannot get out” is known as the greenhouse effect. The greenhouse is shown in fig. (1)



3. GLOBAL WARMING

It is caused by a thick blanket of gases (Carbon dioxide, methane, chlorofluorocarbon, nitrous oxide, etc) and other air pollutants that is building up in the atmosphere become of which the earth get warmer. This blanket of hot house gases permits the entry of sun on the earth but does not permit the return of reflected energy of sun to go back to upper atmosphere. Hence, the earth gets warmer. Infact, the United States emits more carbon dioxide than India, China and Japan.

4. MEASURES TO CURB GLOBAL WARMING

The most important contributor to global warming is the increase in atmosphere CO₂ level due to human activity and OZONE LAYER DEPLETION

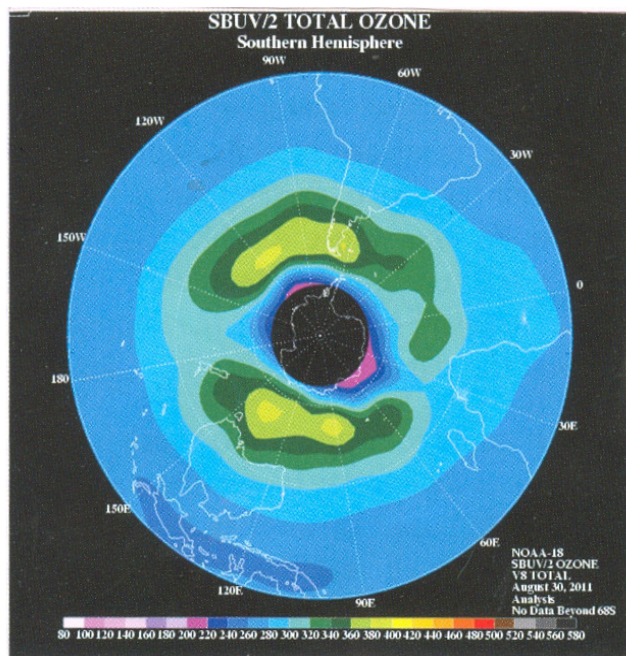
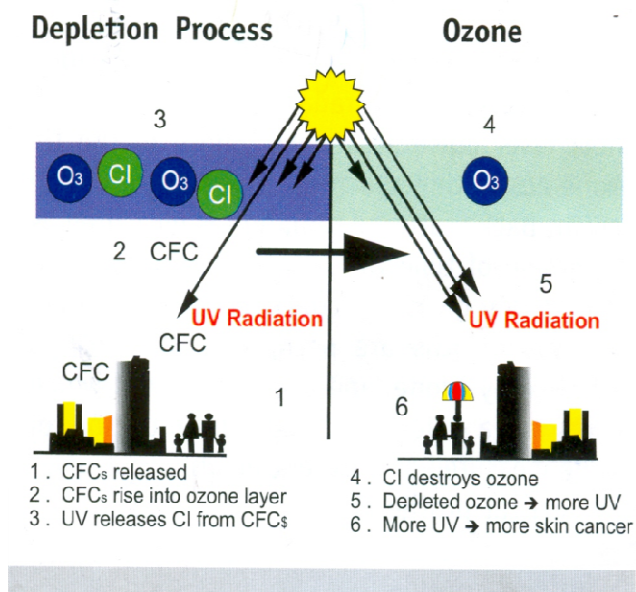


Fig. 3. Evolution of the Antarctic Ozone Hole (1979 – 1987)

OZONE works like an umbrella and protect us from the damaging Ultra-Violet radiation of sun. It filters out UV-B radiations which are biologically harmful and controls the heat budget of the earth. Life on earth's surface would not be

possible without the Ozone Layer. Global Warming can be reduce by:

1. Reducing use of fossil fuels

Utilising renewable resources such as wind, solar and hydropower. Reducing fossil fuel use will also reduce emission of methane, nitrous oxide and ozone.

2. Forestation

CO₂ is consumed by plants and trees. Hence, by reversing deforestation and implementing reforestation programs. CO₂ lends may be reduce in the atmosphere.

3. Individual Efforts

There are many simple steps an individual can take to check AGW – Anthropogenic Global Warming.

4. "Being OZONE Friendly"

Products made with or containing OD's such as CFCs, CTC, CHFCs, halons, methyl chloroform and methyl bromide can contribute to Ozone layer depletion.

5. Be an Ozone friendly consumer

Buy products (aerosol spray cans, refrigerators, fire extinguishers etc.) that are labeled "Ozone friendly" or "CFC free". The product labels should indicate that they do not contain ODSs such as CFCs or halons.

6. Be an Ozone friendly homeowner

Dispose of old refrigerators and appliances responsibly. Consider purchasing new fire exiting wishers that do not contain halon (e.g. dry powder) as recommended by your fire protection authority

7. Be an Ozone friendly farmer

If you use methyl bromide for soil fumigation consider switching to effective and safe alternatives that are currently being used in many countries to replace this Ozone damaging pesticide. Consider options such as integrated pest management that do not rely on costly chemical input. If you do not currently use methyl bromide, do not begin to use it now (you will have to get rid of it in the future.)

8. Be an Ozone friendly refrigeration

Servicing technician regularly check and fix leaks before they become a problem. Help start a refrigerant recovery and recycling programme in your area.

9. *Be an Ozone friendly company*

If your product contains ODSs, change your product formulation to use alternative substances that do not destroy the Ozone layer.

10. *Be an Ozone friendly office worker*

Help your company identify which existing equipment (e. g. water, cooler, air conditioners cleaning etc) and what products it buys aerosol sprays foam cushions (mattresses) use ODSs and develop a plan for replacing them with cost effective alternatives become an environmental leader within your office.

11. *Be an Ozone friendly Teacher*

Inform your students about the importance of protecting the environment and in particular the ozone layer. Teach students about the damaging impact of ODSs on the atmosphere, health impacts and what steps are being taken internationally and nationally to solve this problem. Encourage your students to spread the message to their families.

12. *Be an Ozone friendly community organizer*

Inform your family, neighbor and friends about the need to protect the Ozone layer and help them get

involved and start giving information in your city, town or village.

13. *Be an Ozone friendly citizen*

Read and learn more about effects of Ozone depletion on people, animals and the environment. You can get involved to improve the condition of atmosphere or an individual level. Best solution of global warming is the preservation of the Ozone layer.

5. CONCLUSION

If our environment will be healthy than all activities of human being will go smoothly like communication, networking and signal transmission. All activities related to human being. The ability to observe the things from above the earth, the artificial, satellite rightly called “eyes in the sky”. So we have to save our environment at every possible manner.

REFERENCES

- [1] Journal of Atmospheric and Solar-Terrestrial Physics 63, 1043–47
- [2] R., Williams, E.R., 1999. Dynamics of global thunderstorm activity.
- [3] Huang W G, Gu S F, Gong J C. atmosphere modulation.
- [4] Chen Z Y, Xia M Y. Estimation of current in atmosphere.
- [5] Xu J S, Bao Z T, Liang B X. characteristics of atmosphere.

A Novel Geometry of Wideband Microstrip Patch Antenna with Finite Ground Plane

Sanyog Rawat¹, K. K. Sharma²

^{1,2}Malaviya National Institute of Technology, Jaipur
¹sanyog.rawat@gmail.com

Abstract: In this paper a novel geometry of patch antenna is proposed with improved bandwidth as compared to other antennas of same size. The effect of change in the physical dimension of the ground plane of the proposed antenna is also investigated. The design equations of the antenna obtained by polynomial fitting of the simulation results are also presented. Results show that by selecting suitable ground-plane dimensions and notch area, the impedance bandwidth can be enhanced upto 44.5 %.

1. INTRODUCTION

Microstrip antennas (MSA) have become increasingly popular for microwave and millimeter wave applications, because they offer several advantages over conventional microwave antennas. These advantages include robust structure, easy to fabricate, small size, availability in various shapes, lightweight and conformability with the hosting surfaces of automobiles, aircraft, missiles and direct integration with the microelectronics [1] [2]. MSA, in general consists of radiating conducting patch, a conducting ground plane, a dielectric substrate sandwiched between the two, and a feed connected to the patch through the substrate [3].

The demand for small patch antennas is still growing. Usually the requirement for a small antenna is associated with a reduction in ground plane size to the extent that antenna performance becomes strongly dependent on the ground plane dimensions and position. The effects of change in finite ground plane dimensions on antenna impedance have been investigated in [4] [5]. The bandwidth enhancement upto 16% for a dual polarized antenna is also reported [6]. The effect of different shapes of ground structure on polarization and cross polarization radiation is also investigated in past [7][8].

In this paper, we present novel patch antenna geometry for improvement in bandwidth by truncating the dimensions of finite ground plane. The effects of finite ground plane dimensions of the proposed antenna on different parameters such as efficiency, bandwidth, and gain are also investigated. We have used Method of Moments (MOM) for the analysis of proposed antenna although some other methods such as transmission line model, cavity model, Spectral Domain Full

Wave Analysis, Mixed Potential Integral Equation Analysis, Finite-Difference Time-Domain Analysis (FDTD), Finite Element Method (FEM) etc. exist.

Results show that by selecting suitable ground-plane dimensions and notch area, the impedance bandwidth can be enhanced upto 44.5%. The design equations of the proposed antenna obtained by polynomial fitting of the simulation results are also presented.

The rest of the paper is organized as follows. In section II the proposed antenna geometry is discussed. Simulation results and design equations of the proposed antenna structure are given in section III. The conclusions are given in section IV.

2. PROPOSED ANTENNA GEOMETRY

Patch antennas having rectangular and circular geometry, or their minor variations, are generally used. A minor variant of the commonly used rectangular patch antenna having circular sides of radius 14mm from the centre is shown in figure 1. The proposed antenna geometry, as shown in figure 2, is obtained by cutting a notch in the patch and truncating the ground in conventional geometry. This antenna is designed on glass epoxy FR4 substrate having thickness $h = 1.59$ mm, substrate dielectric constant $\epsilon_r = 4.4$, substrate loss tangent $\tan \delta = 0.024$, and relative permeability $\mu_r = 1$. An air stacking of 1 mm is also added between the patch and dielectric and coaxial feed is provided as shown in figure 3.

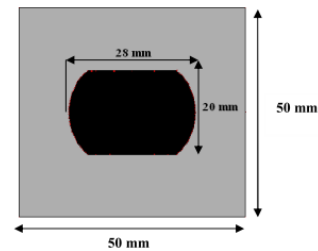


Fig. 1. Geometry of conventional patch antenna on a finite ground plane

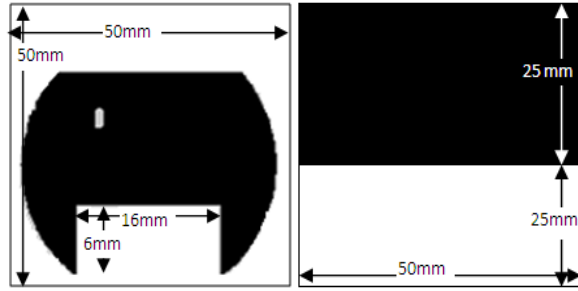


Fig. 2. Proposed geometry of patch antenna with truncated ground plane.

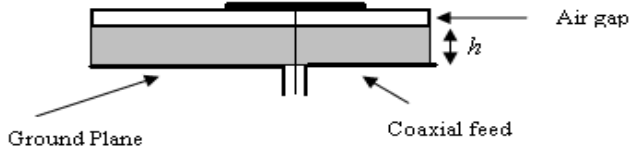


Fig. 3. Side view of the patch antenna with coaxial feed

The proposed antenna acts as a leaky cavity and the field components within the dielectric region of the microstrip antenna may be determined by solving the cavity problem in association with necessary boundary conditions using MOM.

3. SIMULATION RESULTS

The simulations of the proposed antenna geometry are carried out in IE3D software [9] and effects of change of various physical dimensions of the antenna geometry keeping $h=1.59$ mm, substrate dielectric constant $\epsilon_r=4.4$, substrate loss tangent $\tan \delta=0.024$, and relative permeability $\mu_r=1$ constant are investigated.

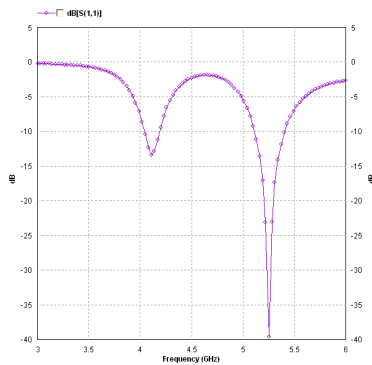


Fig. 4. Simulated variation of return loss with frequency for conventional geometry

The effect of variation of frequency on the reflection coefficient (S_{11}) is investigated for the conventional geometry and is shown in figure 4. The bandwidth can be calculated

from the return loss plot. The bandwidth of the antenna can be said to be those range of frequencies over which the return loss is greater than -10 dB. It is observed from figure 4 that bandwidth of 5.6% is obtained at resonant frequency of 5.25 GHz and -39.62dB return loss.

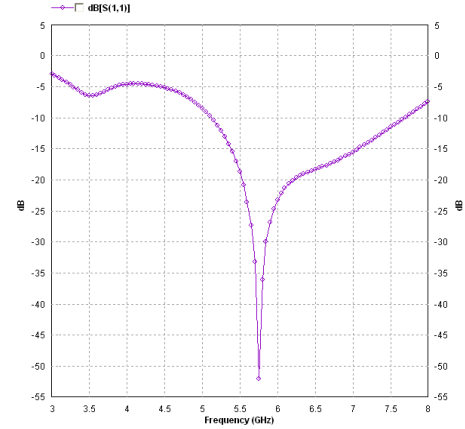


Fig. 5. Simulated variation of return loss with frequency for proposed geometry

The effect of variation of frequency on the return loss (S_{11}) is investigated for proposed patch antenna geometry is shown in figure 5. It is observed that bandwidth of 44.5% is obtained at resonant frequency of 5.75 GHz and -52.11dB return loss.

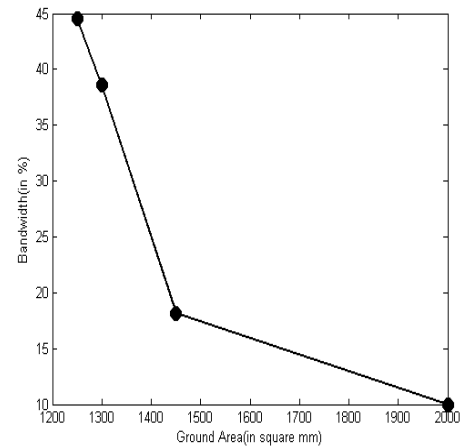


Fig. 6. Simulated variation of bandwidth with ground plane area

The effect of variation of ground plane area A_g on the bandwidth BW is also investigated and the results are shown in figure 6. It can be observed for the given range that bandwidth increases with decrease in ground area.

The relation between bandwidth BW and ground area A_g obtained by polynomial fitting is given by

$$BW = 0.002A_g^2 - 0.5604A_g + 499.03. \quad (1)$$

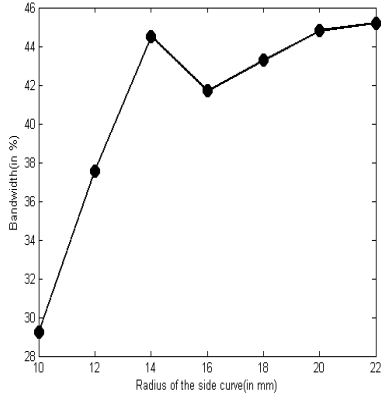


Fig. 7. Simulated variation of bandwidth with radius of side curve in proposed geometry

The effect of variation of radius of side curve in patch r on the bandwidth is also investigated and the results are shown in figure 7. It can be observed for the given range that the bandwidth increases with increase in radius of the side curve.

The relation between bandwidth BW and radius of side curve r obtained by polynomial fitting is given by

$$BW = -0.1725r^2 + 6.6129r - 17.994. \quad (2)$$

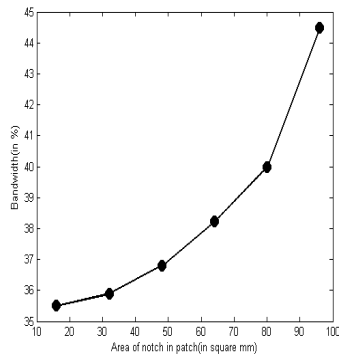


Fig. 8. Simulated variation of bandwidth with notch area in proposed geometry

The effect of variation of notch area in patch A_n on the bandwidth is also investigated and the results are shown in figure 8. It can be observed for the given range that the bandwidth increases with increase in notch area.

The relation between bandwidth BW and notch area A_n obtained by polynomial fitting is given by

$$BW = 0.0017A_n^2 - 0.0828A_n + 36.618. \quad (3)$$

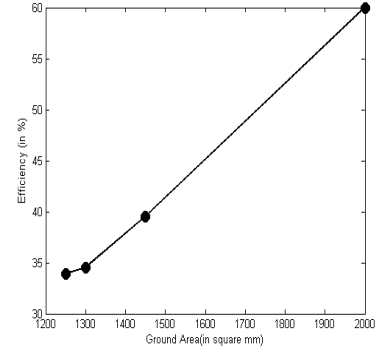


Fig. 9. Simulated variation of efficiency with ground plane area in proposed geometry

The effect of variation of ground area A_g on the efficiency η is also investigated and the results are shown in figure 9. It can be observed for the given range that the efficiency increases with increase in ground area.

The relation between efficiency η and ground area A_g obtained by polynomial fitting is given by

$$\eta = 0.0356A_g - 11.401. \quad (4)$$

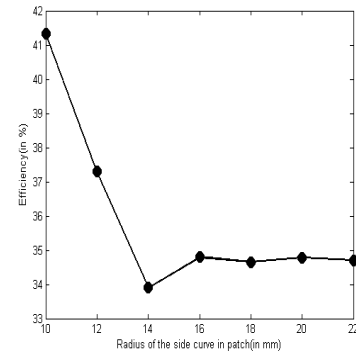


Fig. 10. Simulated variation of efficiency with radius of side curve in proposed geometry

The effect of variation of radius of side curve in patch r on efficiency η is also investigated and the results are shown in figure 10. It can be observed for the given range that initially efficiency decreases with increase in radius of side curve and then becomes constant.

The relation between efficiency η and radius of side curve r obtained by the polynomial fitting is given by

$$\eta = 0.1051r^2 - 3.792r + 68.029. \quad (5)$$

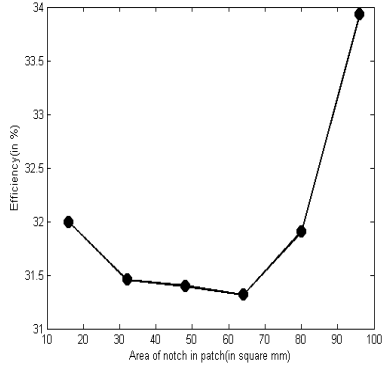


Fig. 11. Simulated variation of efficiency with notch area in proposed geometry

The effect of variation of notch area in patch A_n on efficiency η is also investigated and the results are shown in figure 11. It can be observed from the graph that the efficiency increases with increase in notch area.

The relation between efficiency η and notch area A_n obtained by the polynomial fitting is given by

$$\eta = 0.0011A_n^2 - 0.1008A_n + 33.478. \quad (6)$$

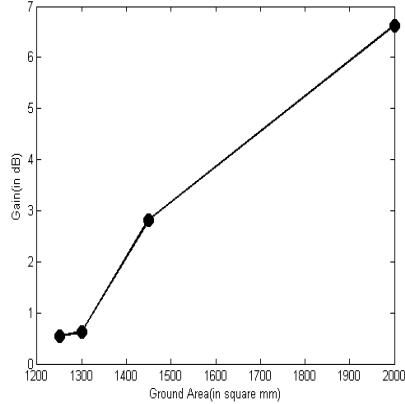


Fig. 12. Simulated variation of gain with ground plane area in proposed geometry

The effect of variation of ground area A_g on gain G is also investigated and the results are shown in figure 12. It can be observed for the given range that the gain increases with increase in ground area.

The relation between gain G and ground area A_g obtained by the polynomial fitting is given by

$$G = 0.0082A_g - 9.6491. \quad (7)$$

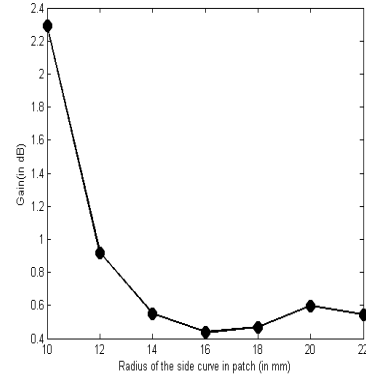


Fig. 13. Simulated variation of gain with radius of side curve in proposed geometry

The effect of variation of radius of side curve in patch r on gain G is also investigated and results are shown in figure 13. It can be observed for the given range that initially gain decreases with increase in radius of side curve and then becomes constant.

The relation between gain G and radius of side curve r obtained by the polynomial fitting is given by

$$G = 0.0279r^2 - 0.998r + 9.22. \quad (8)$$

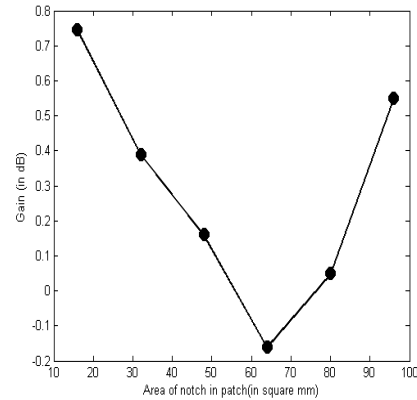


Fig. 14. Variation of gain with notch area in proposed geometry

The effect of variation of notch area A_n on gain G is also investigated and results are shown in figure 14. It can be observed for the given range that the gain initially decreases and then increases with increase in notch area.

The relation between gain G and notch area A_n obtained by the polynomial fitting is given by

$$G = 0.0004A_n^2 - 0.0513A_n + 1.5273. \quad (9)$$

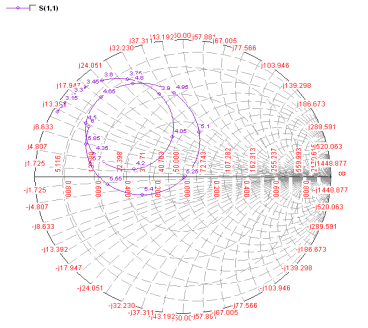


Fig. 15. Variation of input impedance with frequency for conventional geometry

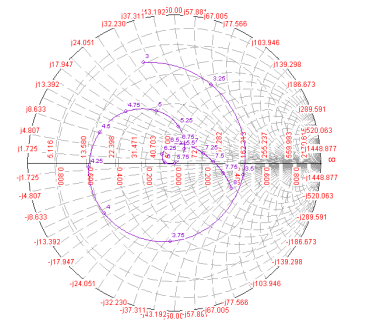


Fig. 16. Variation of input impedance with frequency for proposed geometry

Simulated input impedance variation with frequency is shown in fig. 15 and 16. The circle passes through the centre of the smith chart represents the impedance match of $(49.73-j0.14)\Omega$ for conventional patch and $(50.05-j0.11)\Omega$ for proposed geometry with the coaxial probe and it shows that both geometries have good impedance matching.

4. CONCLUSION

In this paper a novel geometry of patch antenna with finite ground plane is proposed. The simulated results indicate that

proposed patch antenna provides improved bandwidth (44.5) than conventional patch antenna by optimizing the ground plane dimensions. The design equations of the proposed antenna by polynomial fitting are also presented and can be used to obtain desired radiation parameters by substituting the value of physical dimensions.

REFERENCES

- [1] C. A. Balanis, "Antenna Theory Analysis and Design", John Wiley & Sons, Inc., 1997.
- [2] Ramesh Garg, Prakash Bhartia, Inder Bahl, Apisak Ittipiboon, "Microstrip Antenna Design Handbook," Artech House Publications, Boston, London.
- [3] Kin-Lu Wong, "Compact and Broadband, "Microstrip Antennas," Wiley Publication, 2002.
- [4] Kin-Lu Wong and Tzung-Wern Chiou, "Finite Ground Plane Effects on Broad-Band Dual Polarized Patch Antenna Properties" *IEEE Transactions on Antennas and Propagation*, vol. 51, pp.903–904, April 2003.
- [5] A. K. Bhattacharyya, "Effects of Ground Plane Truncation on the Impedance of a Patch Antenna," *IEE Proceedings Microwave Antennas Propagation*, vol. 138, pp.560–564, 1991.
- [6] Takefumi Namiki, Yuichi Murayama and Koichi Ito "Improving Radiation-Pattern Distortion of a Patch Antenna Having a Finite Ground Plane", *IEEE Transactions on Antennas and Propagation*, vol. 51, pp.478–482, March 2003.
- [7] Xue-Xia Yang, Bing-Cheng Shao, Fan Yang, Atef Z. Elsherbeni and Bo Gong, "A Polarization Reconfigurable Patch Antenna With Loop Slots on the Ground Plane", *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 69–72, March 2012
- [8] Chandrakanta Kumar and Debatosh Guha, "Nature of Cross-Polarized Radiations from Probe-Fed Circular Microstrip Antennas and their Suppression Using Different Geometries of Defected Ground Structure (DGS)", *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 1, pp. 92–101, January 2012
- [9] IE3D software, Release 14.65(Zeland Software Inc., Fremont, USA), April 2010.

Stacked Multiband Triangular Fractal Antenna for Mobile Communications

Sumit Kumar¹, Richa Sharma²

¹Department of Electronics and Communication Engineering
Galgotias College of Engineering and Technology, Greater Noida, India

²Assistant Professor, Amity University, Noida, India
¹sumitkumar08@gmail.com, ²s.richa.sharma@gmail.com

Abstract: This paper intends to design a triangular shape fractal patch antenna having resonant frequency of 4.49Ghz. It uses the HFSS for simulation. It is observed that there is an improvement in return losses and VSWR when the iteration is increased and a layer is formed over the antenna. The paper observe that when there is an increase in iteration then the antenna bandwidth increases. Further it is noticed that the second and third iteration shows a multiband behavior in the antenna. The paper uses the substrate material my_mat_ADK with relative permittivity 2.2 . The result of the analysis shows that improved form of antenna is very useful in the cellular communication.

Keywords: Antennas, iteration, multiband behavior

1. INTRODUCTION

The antenna with wider bandwidth and smaller dimension are the requirement of modern telecommunication system. In present scenario there is a fall in the size of electronics system and a miraculous increase in the functionality. However, the antennas have not been changed so far. The wavelength of the antenna, also considered as the distinguished characteristics seems to have influence on the radiation characteristics. But research shows that with the reduction of antenna size there is a change in the bandwidth, gain and efficiency of antenna [1]. The immense progress of wireless industry has sparked at interest in multiband antenna . The most interesting example of a recent multiband antenna development is the incorporation of fractal geometry in to radiator and the Sierpinski gasket antenna. Various other multiband antennas can also be constructed by using fractal geometry. For our analysis in this paper, we are using a triangular shape configuration. A triangular shape fractal patch antenna is separated by a dielectric substrate. However for maximum radiation, low dielectric substrate are preferred

A triangular shape fractal antenna is characterized by it's length, width, i/p impedance gain and radiation pattern. Microstrip triangular patch find various application in design of many useful MIC component such as resonator, circulator and filter[7]. In the present paper, attempts to design

triangular shape fractal antenna of compact size with good radiation and good multiband characteristics. The multi band and ultra wide band properties of antenna are due to their self-similarity of fractal geometry [2]-[3] while the space filling properties [4]-[5] of antenna leads to the miniaturization of antenna.

2. DESIGN SPECIFICATION FOR PROPOSED ANTENNA

The paper observe many parameter using Ansoft HFSS software . It is a triangular shaped fractal patch antenna fabricated on my_mat_ADK substrate with dielectric constant of $\epsilon_r = 2.2$ and a substrate thickness 62mil. The radius of triangular patch is 1.62cm. The size of substrate used is 3cm by 3cm. The assign boundary for patch is perfect E. Here, we are using probe feeding. To calculate resonant frequency of a simple equilateral triangular patch without any degree of sierpinski gasket fractal, the following formula is used[6] :

$$f_{m,n,1} = \frac{2c}{3a(\epsilon_r)^{1/2}}(m^2 + mn + n^2)^{1/2}$$

$$a_{eff} = a + h(\epsilon_r)^{-1/2}$$

$$\epsilon_{eff} = \frac{1}{2}(\epsilon_r + 1) + \frac{1}{4}(\epsilon_r - 1) \left(1 + \frac{12h}{a}\right)^{-1/2}$$

$$f_{m,n} = \frac{2c}{3a_{eff}}(\epsilon_{eff})^{1/2}(m^2 + mn + n^2)^{1/2}$$

where

a = length of equilateral triangular patch h = thickness of substrate

ϵ_r = relative dielectric constant

C = velocity of light

T_{mn} is the resonant fundamental mode hence

m = 0 & n = 1

The calculated side length of equilateral triangular patch from above equation was taken as starting value. For the first

iteration, the paper takes one third area of the patch of antenna. In the second iteration, one third area of first iteration is used. The same procedure is used for the third iteration, in case of third iteration one third of second iteration can be taken

3. RESULT

Three iteration result performed on the triangular patch to get the desired fractal antenna are as follow:

4. RESULT FOR ITERATION 0

When triangular fractal antenna structure having zero iteration has show in fig. On simulating this structure with the help of Anasoft HFSS, the following result were obtained



Fig. 1 (iteration 0)

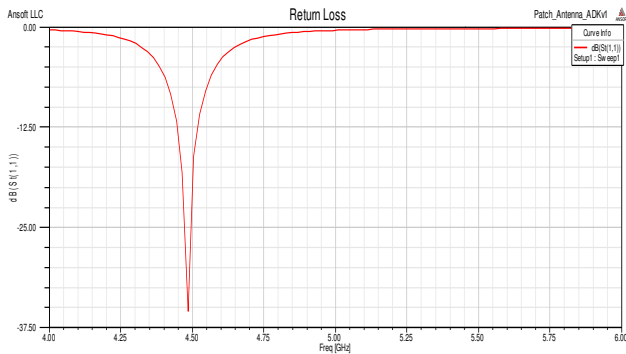


Fig. 2. (Return losses for various frequencies)

That is return loss is found to be -37.12dB at 4.49 GHz frequency

5. RESULT FOR ITERATION 1

The structure of triangular fractal antenna with first iteration is as follows:



Fig. 3 (iteration 1)

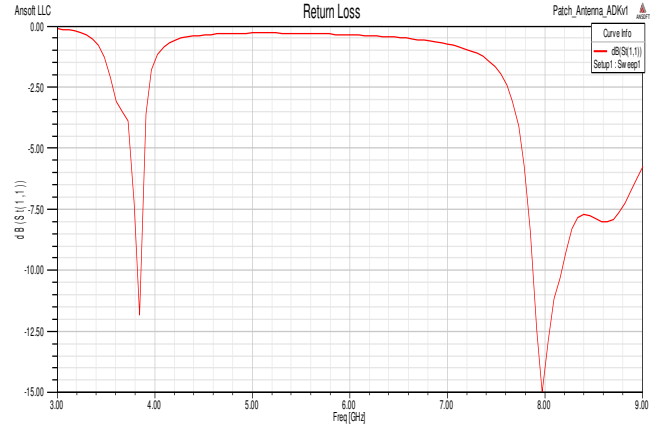


Fig. 4. (Return losses for various frequencies)

That is return loss is found to be -11.80dB at 3.8440 GHz frequency and -15dB at 7.96Ghz frequency.

6. RESULT FOR ITERATION 02

The structure for triangular fractal antenna with 2nd iteration is:

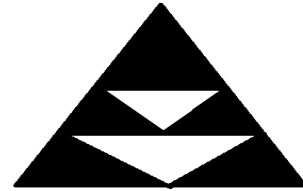


Fig. 5 (iteration 2)

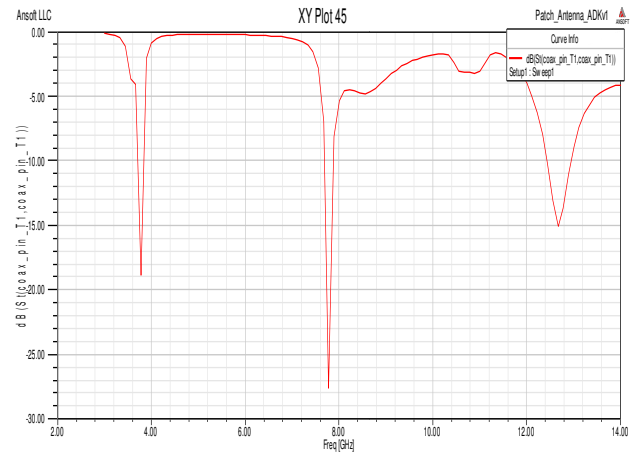


Fig. 6. (Return losses for various frequencies)

That is return loss is found to be -18.82dB at 3.76 GHz frequency and -27.56dB at 7.7697Ghz frequency and -15.04dB at 12.66Ghz frequency .

7. RESULT FOR ITERATION 03

The structure for triangular fractal antenna with 3rd iteration is:



Fig. 7 (iteration 3)

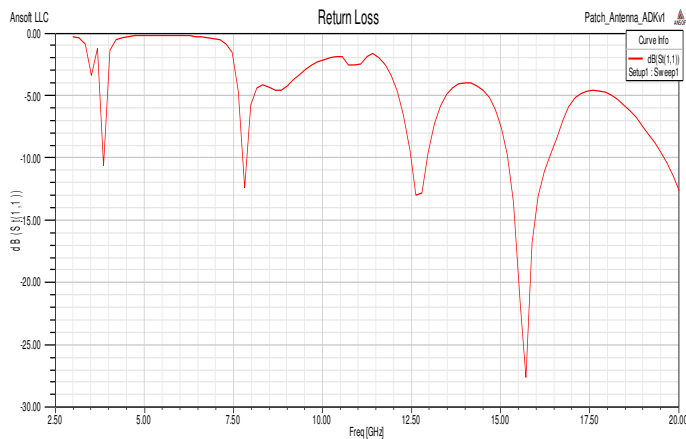


Fig. 8 (Return losses for various frequencies)

That is return loss is found to be -10.42dB at 3.82GHz frequency and -12.18dB at 7.77GHz frequency and -12.94dB at 12.62GHz frequency and -27.56dB at 15.68GHz frequency.

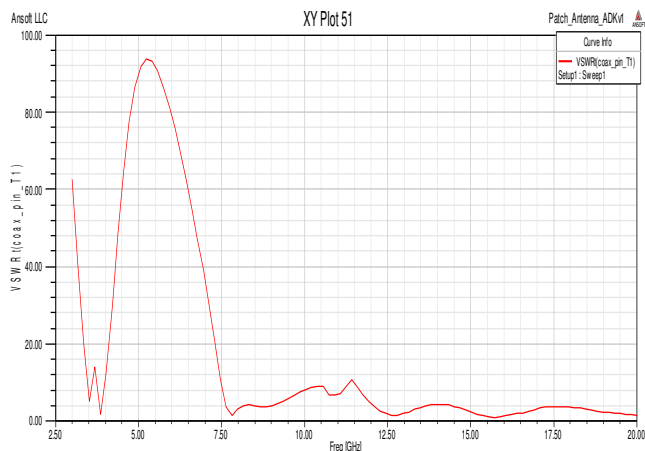


Fig. 9 (VSWR vs. frequency)

8. RESULT FOR AFTER LAYER IN ITERATION 03

The structure for triangular fractal antenna with LAYER in 3rd iteration is:

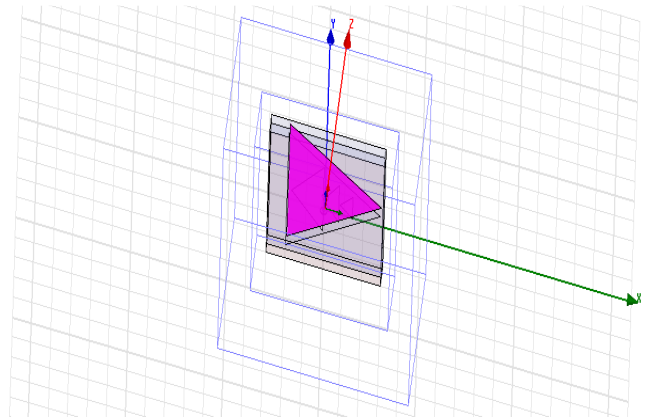


Fig. 10 (after layer)

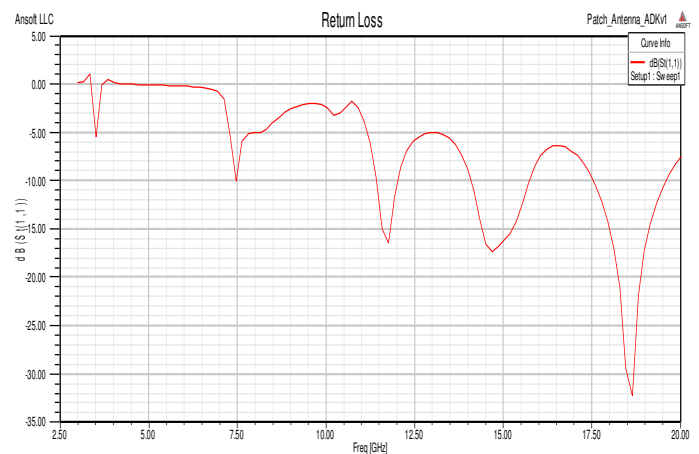
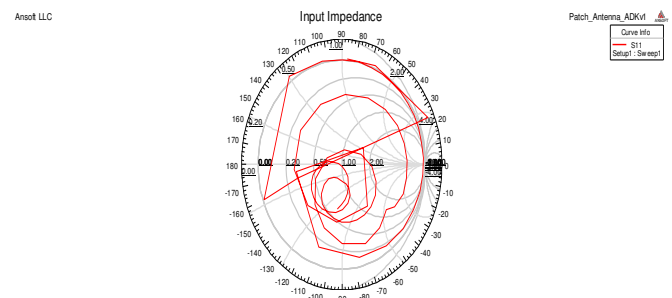


Fig. 11. (Return losses for various frequencies)

That is return loss is found to be -10.04dB at 7.45GHz frequency and -16.40dB at 11.73GHz frequency and -17.40dB at 14.65GHz frequency and -32.19dB at 18.60GHz frequency.

This fig. 12 show the input impedance



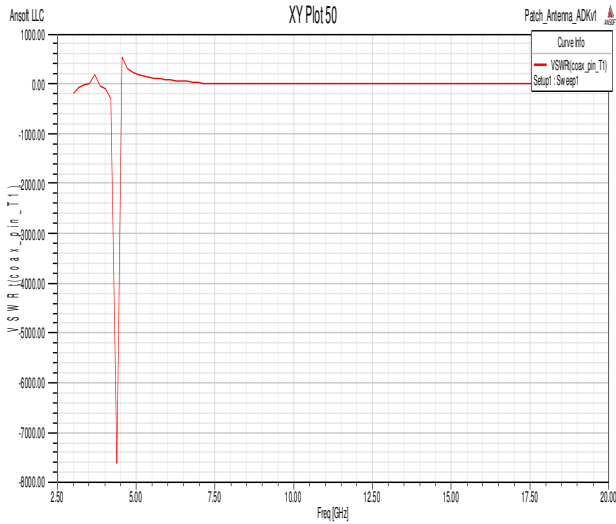


Fig. 13. (VSWR vs frequency)

9. CONCLUSION

Triangular shape fractal antenna up to third iteration has been built & simulated using Anasoft HFSS. It is observed that after third iteration we are getting four band below -10db but after layer also getting four band with good

performance and show the multiband behavior. Thus the result shows that this improved antenna can be highly beneficial in the field of cellular communication.

REFERENCES

- [1] J. Bahl and P. Bhartia, "Microstrip Antennas", Dedham, Ma, Artech. House, 1981
- [2] C. Puente, J. Romeu, R. Pous, X. Garcia, and F. Benitez, "Fractal multiband antenna based on the Sierpinski gasket," *Electron. Lett.*, Vol.32, no.1, PP.1-2 Jan.1996 .
- [3] C. Puente, J. Romeu, R. Pous, and A. Cardama, "On the behavior of the Sierpinski multiband fractal antenna," *IEEE Trans. Antenna Propagate*, Vol.46, PP.517-524, Apr.1998.
- [4] S.N. Khan, J. Hu, J. Xiong, and S. He, "Circulator fractal monopole antenna for low VSWR UWB application", *Progress in Electromagnetics Research Letters*, Vol.1, PP.19-25, 2008 .
- [5] E. Lule, et al, "Koch island fractal ultra wideband dipole antenna", *IEEE, Antenna and Propagation Society International Symposium*, Vol.3, PP.2516-2519, June 2004.
- [6] Dahele, J.S., on the resonant frequencies of the triangular patch antenna, "IEEE Transactions on Antennas and Propagation vol. 35, No.1, 100-101, 1987.
- [7] R. L. Yadava, M. Ram, and S. Das, "Multiband Triangular Fractal Antenna for Mobile Communication", *International Journal of Engineering Science and Technology*, Vol.2(11), 6335-6348, 2010 .

An Enhancement in Data Compression Using H.264 /AVC

Ankita Awasthi¹, Anshika Salaria², Samta Suman Lodhi³

¹Amity Institute of Telecom Technology and Management Amity University (U.P)
ankitaawasthi44@gmail.com

²Amity Institute of Telecom Technology and Management Amity University (U.P)
anshikasalaria3005@gmail.com

³2709 West Royal Lane, Irving, Texas (U.S.A), samtalodhi@gmail.com

Abstract:-The ever increasing bandwidth requirements for transmission of video signals in mobile and internet environment has necessitated video compression and attempts to compare the low bit rate characteristics of the major video compression methods. This paper makes use of H.264 using DCT and wavelet based video compression. Also attempts are made to compare the results of these methods. Initially the compressed signal / data is transmitted and the receiving end the video signals are reconstructed.

Keywords: H.264, DCT, Video Compression, Wavelet, Low Bit Rate, Matlab, etc.

1. INTRODUCTION

Despite rapid progress in mass-storage density, processor speeds, and digital communication system performance, demand for data storage capacity and data-transmission bandwidth continues to outstrip the capabilities of available technologies. The recent growth of data intensive multimedia-based web applications have not only sustained the need for more efficient ways to encode signals and images but have made compression of such signals central to storage and communication technology.

Compression is useful because it helps reduce the consumption of expensive resources, such as hard disk space or transmission bandwidth. On the downside, compressed data must be decompressed to be used, and this extra processing may be detrimental to some applications. Compressed video can effectively reduce the bandwidth required to transmit video via terrestrial broadcast, via cable TV, or via satellite TV services. Most video compression is lossy it operates on the premise that much of the data present before compression is not necessary for achieving good perceptual quality. Video is basically a three-dimensional array of color pixels. Two dimensions serve as spatial (horizontal and vertical) directions of the moving pictures, and one dimension represents the time domain. A data frame is a set of all pixels that correspond to a single time moment. Basically, a frame is the same as a still picture. The performance of H.264 generally degrades at low bit-rates

mainly because of the underlying block-based Discrete Cosine Transform (DCT) scheme. More recently, the wavelet transform has emerged as a cutting edge technology, within the field of image & video compression. Many efforts have been taken in past to discuss image compression techniques [3] [6].

2. PROBLEM STATEMENT

Discrete Cosine Transformation is mainly used for image, video compression but it has several disadvantages such as

- Only spatial correlation of the pixels inside the single 2-D block is considered and the correlation from the pixels of the neighboring blocks is neglected.
- Undesirable blocking artifacts affect the reconstructed images or video frames. (High compression ratios or very low bit rates).
- DCT function is fixed. i.e. it cannot be adapted to source data
- Does not perform efficiently for binary images (fax or pictures of fingerprints) characterized by large periods of constant amplitude, followed by brief periods of sharp transitions.

So, because of all this reasons DCT does not provide efficient image/video compression and we may get noisy image while decompressing data.

3. WAVELET

Wavelet can often compress or de-noise a signal without appreciable degradation. Wavelet transforms are broadly divided into three classes: the continuous wavelet transform, the discretized wavelet transform and multi-resolution based wavelet transform. DWT is good for signal having high frequency components for short durations and low frequency components for long duration e.g. images. When a wavelet transform of the image is performed, a coefficient in a low sub-band can be thought of having four descendants in the

next higher sub-band. The four descendants each have four descendants in the next higher sub-band. Discrete wavelet transform (DWT), transforms a discrete time signal to a discrete wavelet representation [1] [4]. A 2D wavelet transforms works as follows [5]:

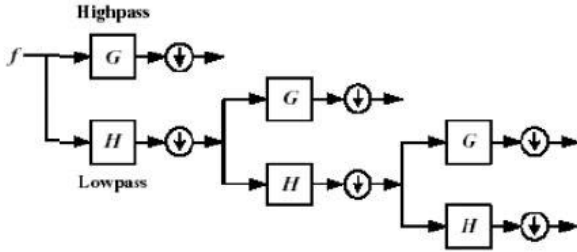


Fig. 1. Wavelet Decomposition

Wavelets are functions defined over a finite interval and having an average value of zero. In wavelet decomposition, an image is decomposed into four components namely approximate coefficients, Horizontal coefficients, diagonal coefficients and vertical coefficients. This kind of two-dimensional DWT leads to a decomposition of approximation coefficients at level j in four components: the approximation at level $j+1$, and the details in three orientations (horizontal, vertical, and diagonal). Many research paper on performance analysis has been discussed in past using wavelets transform [2]. The following chart describes the basic decomposition step for images

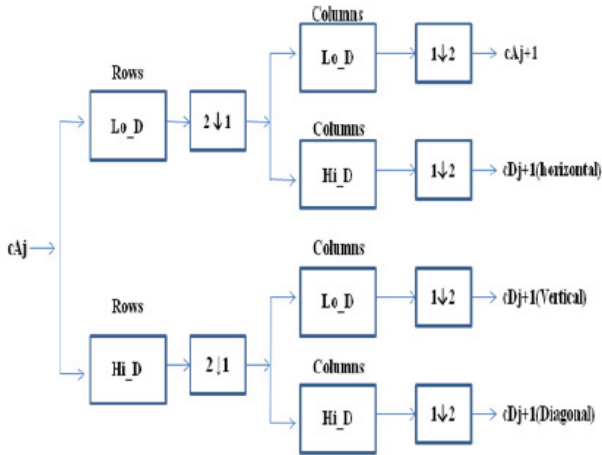


Fig. 2. Decomposition Steps

Algorithm for Image decomposition using DWT

- Apply low pass and high pass on an image.
- Perform column wise down sampling on resultant values of step 1.

- Repeat step 1 on the output of step 2.
- Apply row wise down sampling.(now we get Approximate coefficients, horizontal details, diagonal details and vertical details)
- Stop if the stopping criterion is met or apply step 1 on approximate coefficient.
- For down sampling we can either use averaging method with details or pure interpolation method.
- To reconstruct an image apply the reverse technique of the above algorithm.

B. Advantages of DWT

- Allows good localization both in time and spatial frequency domain. Transformation of the whole image.Higher compression ratio.
- Higher flexibility.
- Better Performance.

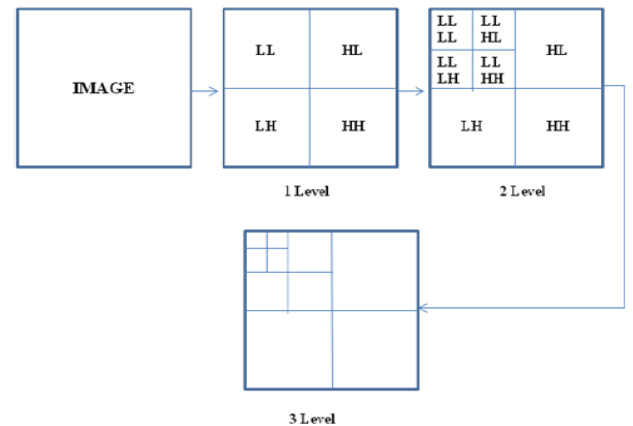


Fig. 3. Image decomposition using wavelets

4. H.264/ AVC

H.264 / AVC has a variety of new features that improve the picture quality and compression such as:

A. In loop deblocking

1. Deblocking is a CPU Intensive technique that attempts to remove blocking artifacts in the decoded picture.
2. H.264 enforces deblocking on every frame during both encoding and decoding.
3. The result is that encoding becomes more efficient because there is less noise present in reference pictures but consequently there is no option to disable deblocking to boost playback performance on slower systems.

B. Quarter- pixel motion estimation

H.264 use quarter pixel precision for motion search and this leads to longer search times during encoding as well as more complex texture reconstruction during playback.

5. CONCLUSION

H.264 is a next-generation video compression format. H.264 is also known as MPEG-4 AVC. Developed for use in high definition systems such as HDTV, Blu-ray and HD DVD as well as low resolution portable devices such as Sony's PSP and Apple's iPod, H.264 offers better quality at lower file sizes than both MPEG-2 and MPEG-4 ASP. The standardization of the first version of H.264/AVC was completed in May 2003. The H.264 standard can be viewed as a "family of standards", the members of which are the profiles described below. A specific decoder decodes at least one, but not necessarily all profiles. The decoder specification describes which of the profiles can be decoded. It has certain advantages such as an in-loop deblocking filter that helps prevent the blocking artifacts common to other DCT-based image compression techniques, resulting in better visual appearance and compression efficiency. A quantization design including Logarithmic step size control

for easier bit rate management by encoders and simplified inverse-quantization scaling Frequency-customized quantization scaling matrices selected by the encoder for perceptual-based quantization optimization.

REFERENCES

- [1] Antonini M, Barland M, Mathieu P, Daubechies I. "Image coding using the wavelet transform". IEEE transactions on Image Processing 1992;2:205-20.
- [2] Grgic S, Grgic M, Zovko-Cihlar B. "Performance analysis of image compression using wavlets" .IEEE Transaction on Industrial Electronics 2002;48:682-95.
- [3] Grgic S, Kers K, Grgic M. Image compression using wavlets. Proceedings of the IEEE international symposium on industrial electronics, ISIE'99, Bled, Slovenia, 12-16 July 1999, p.99-104.
- [4] Hilton ML, Jawerth BO, Sengupta A. Compressing still and moving images with wavelet Multimedia Systems 1994;2:218-27.
- [5] Lewis AS, Knowles G. Image compression using the 2-D wavelet transform. IEEE Transactions on image Processing 1992; 2:244-50.
- [6] Taubman D, Marcellin MW. JPEG2000 image compression: fundamentals, standards and practice. Dordrecht: Kluwer Academic Publishers; 2002

Study of Smart Antennas and their use in Wireless Communication Systems

Amritpal Singh Bhinder¹, Rajat Singh²

¹Amity Institute of Telecom Technology & Management, Amity University, bhinderamritpal@yahoo.com

²Amity School of Engineering, Amity University, shanks_rajat@yahoo.in

Abstract: The use of smart antennas in mobile radio communications cellular networks such as GSM network, to mitigate the fading effects and to increase the traffic capacity by exploiting the Spatial Division Multiple Access (SDMA) is actually of great interest. Although the implementation of smart antennas is mainly investigated on base stations, due in particular to obvious cumbersomeness reasons, it is still interesting to evaluate the smart antennas performance at the mobile level. The most important feature of a smart antenna is its beam forming capability. During beam forming the smart antenna creates a directional beam toward the desired user and nulls the signal in the directions of undesired users by appropriately adjusting the magnitude and phase of the signal transmitted by each of its elements. In comparison to Omni-directional transmissions, beam forming reduces interference, allowing more concurrent transmissions in the network. Moreover, by concentrating the transmission energy in a specific direction, beam forming creates a signal that is in order of the magnitude stronger than that of the signals in other directions. This technique can be used to increase the coverage of a particular area or data rate or the spectral efficiency of the system. In this paper different types of techniques used for the working of smart antennas. The amplitude or time delay (phase) of the signals received by all the antennas are modified then combined in such a manner as to improve reception of the desired signal.

Keywords: Smart Antenna, Beam forming, SDMA

1. INTRODUCTION

A smart antenna is an array antenna composed of two or more antennas. The amplitude and/or time delay (phase) of the signals received by all the antennas are modified then combined in such a manner as to improve reception of the desired signal. Moreover, by concentrating the transmission energy in a specific direction, beam forming creates a signal that is orders of magnitude stronger than that of the signals in other directions. This technique can be used to increase the coverage of a particular data rate or the spectral efficiency of the system. The increased signal-to-noise ratio results in a larger gain in the direction of the user, and also provides better control of the distribution of spatial interference in the cell. Beam forming can be applied to the downlink and uplink. Smart antennas have gained great interest among researchers during recent years. Wireless operations are currently searching for new technologies to be

implemented into the existing wireless communications infrastructures for capacity enhancement and quality improvement. Such research efforts will enable wireless carriers to boost the spectral efficiency of their networks so as to meet the explosive growth of wireless communications industry and take advantage of the huge market opportunity. Deployed at the base station of the existing wireless infrastructures, smart antennas are capable of bringing outstanding capacity improvement [1].

Until now, the investigation of smart antennas suitable for wireless communication systems has involved primarily uniform linear arrays (ULA). Different algorithms have been proposed for the estimation of the direction of arrivals (DOAs) of signals arriving to the array and several adaptive techniques have been examined for the shaping of the radiation pattern under different constraints imposed by the wireless environment. Furthermore, in the literature for adaptive antennas so far, little attention has been paid to other array topologies. Two key components of smart antenna technology examined here are direction of arrival (DOA) estimation and adaptive beamforming. With the former, it is feasible to determine the angles from which sources transmit signals towards an antenna array. With the latter, an antenna radiation pattern beam maximum can be simultaneously placed towards the intended user and ideally nulls can be placed towards directions of interfering signals.

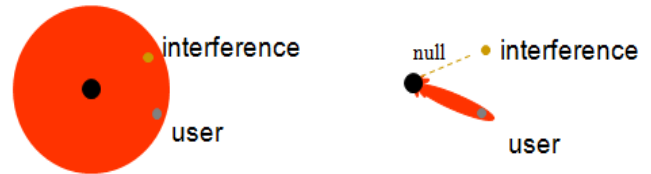


Fig. 1. Beamforming pattern of smart antenna and omni-directional

We consider the MUSIC algorithm for DOA estimation. Adaptive beamforming is achieved using the RLS algorithm.

2. THEORY

There are two basic types of smart antennas. As shown in Fig. 2, the first type is the phased array or multibeam

antenna, which consists of either a number of fixed beams with one beam turned on towards the desired signal or a single beam (formed by phase adjustment only) that is steered toward the desired signal. The other type is the adaptive antenna array as shown in Fig. 3, which is an array of multiple antenna elements, with the received signals weighted and combined to maximize the desired signal to interference plus noise power ratio[2]. This essentially puts a main beam in the direction of the desired signal and nulls in the direction of the interference. A smart antenna is therefore a phased or adaptive array that adjusts to the environment. That is, for the adaptive array, the beam pattern changes as the desired user and the interference move; and for the phased array the beam is steered or different beams are selected as the desired user moves[3].

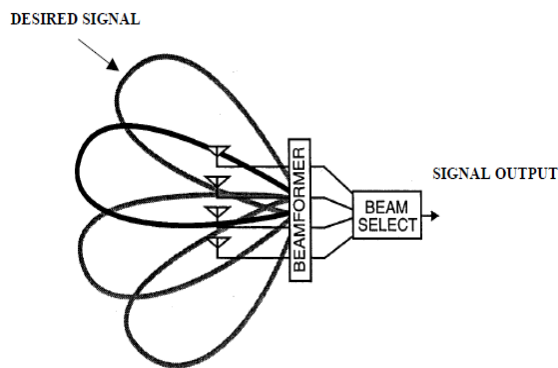


Fig. 2. Phased Array

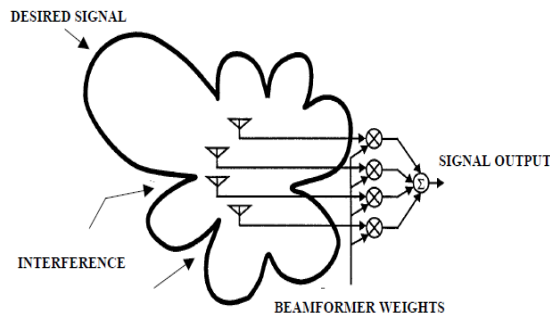


Fig. 3. Adaptive Array

The gain of a smart antenna is normally greater than that of an Omni-directional antenna. Also, when compared to an omni-directional antenna, smart antenna has higher reach ability i.e., a larger directional range. Beam forming is the term used to define the application of weights to the inputs of an array of antennas to steer the reception of the antenna array in a particular direction, called the look direction or the main lobe[4]. Beam forming techniques aims at enhancing the captured sound quality by using the diversity in the received signals of the microphone array depending on the location of the source and the hindrance. The two significant functions of smart antennas are Direction of Arrival and

Adaptive Beam forming [5]. Adaptive beam forming systems uses an adaptive array processing for the creation of nulls in the direction of interference as well as powerful beams in the direction of desired user

3. RESULTS

Smart Antenna Systems use the additional degrees of freedom offered by their multiple antennas to exploit, among other things, multipath in the propagation environment. Therefore, by construction, antenna design of smart antenna systems cannot be assessed by simple performance metrics such as gain, polarization and efficiency alone. At a minimum, performance has to be considered in the context of the nature and degree of the multipath. Capacity, the maximum possible throughput, is an appropriate performance metric when the antennas are properly combined with their propagation environment but nothing more is known about the system.

When, additionally, the specific Link and Media Access Control (MAC) layer characteristics of the system are taken into account, the actual throughput of the communication link becomes a more appropriate performance metric. Therefore, Design Requirements for Smart Antenna Systems must be considered under a cross layered optimization umbrella. A Cross-Layered design approach of Multiple Input Multiple Output (MIMO) antenna systems is presented in this talk. An electromagnetic exact formulation from baseband-to-baseband of a Smart Antenna System is given. The formulation consists of full wave analyses of the antenna arrays involved on both sides of the link and plane wave decomposition for the propagation environment. Subsequently, the baseband signals are fed into link simulators, specific for each system of interest, to provide estimates of the Bit Error Rate (BER) and throughput. Calibration and Channel estimation algorithms are described for Time Division Duplex (TDD) systems, such as the IEEE 802.16 (WiMAX). In particular, a design requirement to ensure reciprocity calibration of a TDD smart antenna based system is identified. The state of the art in designing antennas for terminals and for base stations is outlined.

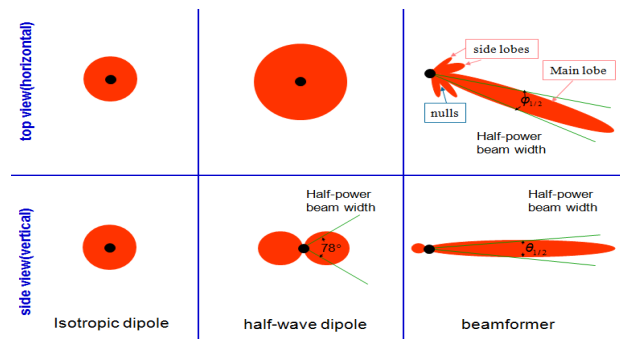


Fig. 4. Beam pattern of different types of beamforming antennas

4. CONCLUSION

The uses of smart antenna are now not only constrained in the communication networks but also in the medical field as well as other application are there like detection, tracking etc. Smart antenna is a new buzzword for adaptive antennas. Since commercial communications technology is driving the need for this type of antennas, there is a different niche from the military drive for adaptive antennas. In any event, the explosion of wireless communications applications necessitates the need for antennas that optimally adjust their performance based upon the signal environment. The coming time is very bright for the smart antennas in every field.

The successful adoption of smart antennas relies on considering the particular features of the technology at an early stage in the design of future systems. In this context the major trends in the area of smart antennas, such as reconfigurability to varying channel propagation and network conditions, cross-layer optimization, and multi-user diversity techniques, have been discussed. Moreover, challenges such as the design of a suitable simulation

methodology and the accurate modeling of channel characteristics, interference, and implementation losses have been presented along with market trends, future projections, and the expected financial impact of smart antenna systems deployment.

REFERENCES

- [1] Ioannides, P.; Balanis, C.A., "Uniform Circular for Smart Antennas," IEEE Trans. Antennas and Propagation, vol. 47, pp. 192 -206, Aug.2005.
- [2] A. Alexiou and M. Haardt, "Smart antenna technologies for future wireless systems: Trends and challenges", IEEE Commun. Mag., vol. 42, no. 9, pp. 90-97, Sep. 2004.
- [3] Jin Yong-Hong, Geng Jun-Ping, Fan Yu. Smart Antenna In Wireless Communication [M] Beijing Post and Telecommunications University Press.2006.
- [4] J.H. Winters and M.J. Gans, "The Range Increase of Adaptive versus Phased Arrays in Mobile Radio Systems," IEEE Trans. Vehicular Technology, vol. 48, no. 2, pp. 353-362, Mar. 1999.
- [5] Lal C.Godara, "Application of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations", *Proceedings of the IEEE*, Vol. 85, No. 8, pp. 1195- 1245, Aug 1997.

Parametric Analysis of Co-axial Probe fed Rectangular Dielectric Resonator Antenna

Neeraj Kumar¹, Arvind Kumar²

^{1,2}Amity Institute of Telecom Technology and Management, Amity University, Noida (UP), India
¹neeraj767@gmail.com, ²arvind9356@gmail.com

Abstract: This paper includes design of a Rectangular Dielectric Resonator Antenna (RDRA). Antenna is excited using co-axial probe. A detailed parametric study of probe pin insertion into the volumetric radiating source is carried out to analyze the characteristics of the proposed antenna. DR, a radiating source improves the radiation power factor of the antenna. The theoretical performance of the antenna is verified by full-wave simulations and experimental data obtained from a prototype at frequency 3.68 GHz. The antenna provides a -10 dB fractional impedance on the bandwidth of 20 MHz and, the specified frequency range of 3.6 GHz- 3.8GHz. The VSWR value of antenna is maintained in the range of 2-3 at the frequency of operation. Gain of prototype is 2.3 dBi.

Keywords: Rectangular Dielectric Resonator Antenna; Co-axial Probe; Impedance; Bandwidth; VSWR; Gain.

1. INTRODUCTION

Today there is demand of high data rate for efficient wireless communication. Modern Antenna should be capable to handle this huge demand of Bandwidth without compromising with the antenna characteristics such as VSWR, Gain, radiation Pattern and many more. It can be achieved by operating the wireless mobile systems at the millimeter wave frequencies [1-2]. Since 1991, Dielectric Resonator Antenna (DRA) are being preferred over the conventional micro-strip antennas because of its features such as small size, low cost, good bandwidth, high gain, Low ohmic loss and light weight. The DRA can be used at millimetre frequency bands and they are available in basic shapes such as rectangular, cylindrical, spherical and hemispherical geometries.

The designed antenna is rectangular in shape as it offers more design flexibility since two of the three of its dimensions can be varied independently for fixed resonant frequency and known dielectric constant of the material, allowing greater degree of freedom.

A feed-line is used to excite to radiate by direct or indirect contact. There are many different techniques of feeding and four most popular techniques are coaxial probe feed, micro-

strip line, aperture coupling and proximity coupling [3-4]. Proposed Antenna is excited using coaxial probe fed technique. Coaxial probe feeding is feeding method in which the inner conductor of the coaxial is attached to the radiating element of the antenna while the outer conductor is connected to the ground plane.

2. CALCULATIONS

Theoretical calculation has been performed using MATLAB software, where the mathematical equations were implemented and solved to get the desired dimensions of antenna radiating element for the $TE_{\delta 11}^x$ mode, resonant frequency of 2.5 GHz, VSWR of 2, relative permittivity of resonator as 15. Minimum to maximum value of w/h and d/h is $1 \leq (\text{ratio}) \leq 3$, where ratio is w/h or d/h. Result obtained for the selected mode is presented in Table I.

Table I: Dimension of Rdra Obtained on Matlab

W (cm)	D (cm)	H (cm)	Q-factor	BW (%)
1.784	5.352	1.784	6.93	10.2

3. PROPOSED DESIGN

The proposed design uses a substrate of material Rogers RT/Duroid 5880 (tm) having relative permittivity of 2.2 and dimension 10 cm X 9 cm X .32 cm. The upper surface of the substrate has finite conductivity layer. This has been done to minimize the backlobe radiation phenomenon. The rectangular dielectric resonator of relative permittivity 15 is used, having dimension of 5.352 cm X 1.784 cm X 1.784 cm. The co-axial probe feed mechanism is used for the excitation of antenna. In Fig. 2(a), the proposed antenna is presented in 3D view. The parametric study of the antenna is done by varying the probe penetration length, l into the dielectric material of the resonator. The variation of the length is done from 0.5111 cm to 0.8111 cm in step size of 0.111 cm. The antenna designed for parametric study, showing probe length l, is shown in Fig. 2(b).

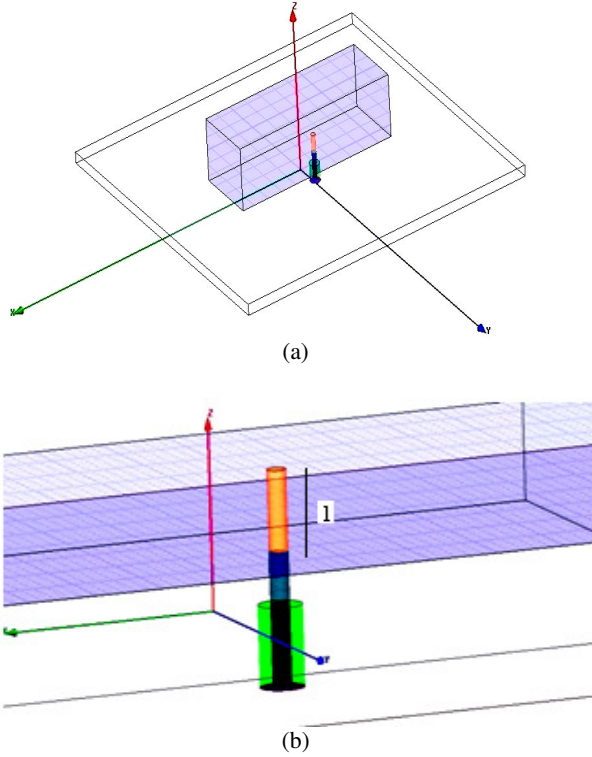


Fig. 2(a) 3D View of RDRA Antenna, (b) Parametric variation of Probe penetration inside the antenna as Project Variable, l

4. SIMULATION RESULTS

The designed antenna is simulated on Ansoft HFSSv13 simulation software. The parametric variation of the project variable, l was done. The result obtained was studied in detail to know the optimum behavior of RDRA.

A. Parametric Study of RDRA Antenna by Varying the Probe Length, l Penetration into the Dielectric Resonator.

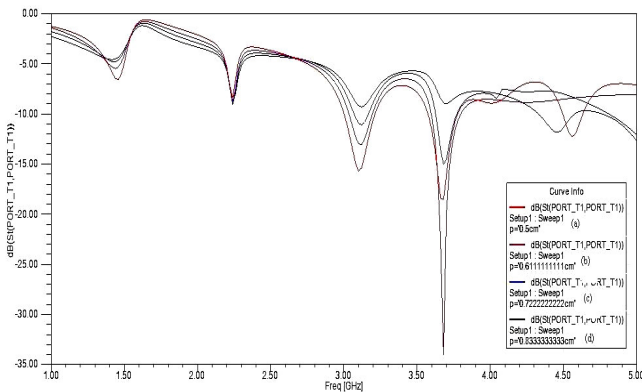


Fig. 2(a) S_{11} Vs Frequency Plot of RDRA Antenna graph showing parametric variation of Probe length, l (a) $l = .5111$ cm, (b) $l = .6111$ cm (c) $l = 0.7111$ cm (d) $l = 0.8111$ cm

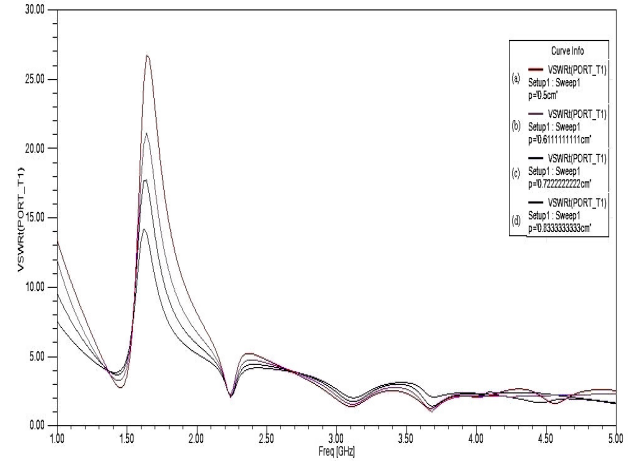


Fig. 2(b) VSWR Vs Frequency Plot of RDRA Antenna graph showing parametric variation of Probe length, l (a) $l = .5111$ cm, (b) $l = .6111$ cm (c) $l = 0.7111$ cm (d) $l = 0.8111$ cm

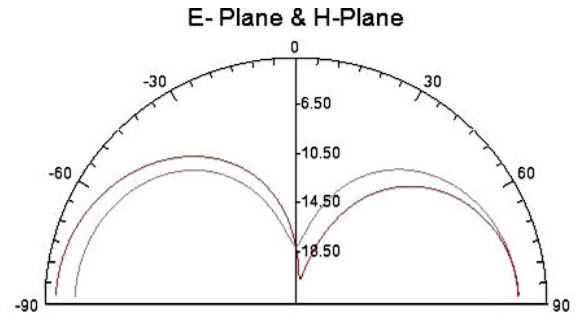


Fig. 2(c) E and H-Field Pattern of RDRA

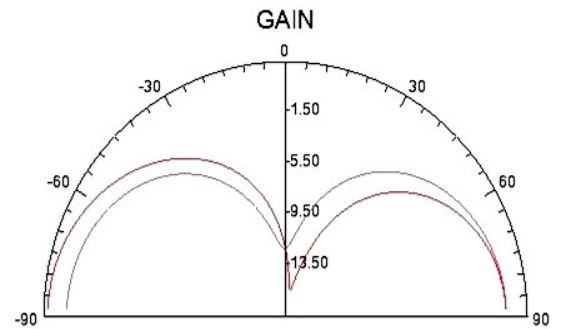


Fig. 2(d) Gain of RDRA in terms of radiation Pattern

5. DISCUSSION

As discussed in Section III, the parametric variation in the probe length, l was done by setting, l as project variable. The simulated results are shown in Figure 2. Return Loss curve and VSWR curve of proposed antenna has been presented in Figure 2 (a), and (b) respectively. Figure 2 (c) and (d) depicts E Field and H-Fields of RDRA in terms of Radiation Pattern. During the variation of probe length, l the matching

frequency gets shifted to higher value but the fractional bandwidth of the antenna is increased.

The S_{11} Vs Frequency plot of antenna shows that antenna is well matched at resonant frequency of 3.68 GHz, having return loss of -34.02 dB. Antenna offers -10 dB bandwidth of interval within frequency range 3.6 GHz – 3.8 GHz. Antenna provides impedance bandwidth of 20 MHz. This is the optimized design for probe penetration to 0.6111 cm inside the radiating material.

From Fig. 2(b), VSWR value of the antenna is found to be in between 2-4 for the matching frequency of the RDRA antenna, which is advantageous. Fig. 2 (c) shows the E and H- field pattern of the antenna. Gain of the antenna in terms of radiation pattern has been presented in Fig. 2 (d), which is found to be 1.72.

6. CONCLUSION

A Compact Rectangular Dielectric Resonator Antenna has been proposed. Antenna is excited using co-axial probe which has advantage of easy fabrication; easy to match; low spurious radiation. A detailed parametric study of probe pin insertion into the volumetric radiating source is carried out to analyze the characteristics of the antenna. Ratio of probe pin insertion length, l ; to height of radiating source, h ; has been obtained 0.34. With proper design it is observed that the

resonance of the probe and that of the dielectric structure itself may be merged to achieve extremely wide bandwidth over which the antenna polarization and radiation pattern are preserved. The proposed RDRA has resonant frequency of 3.68 GHz and offers bandwidth of 20 MHz having maximum return loss of -34 dBi, indicating its suitability for using as indoor/outdoor wireless applications.

7. ACKNOWLEDGEMENT

We are very much thankful Amity Institute of Telecom Technology & Management, Amity University for providing essential lab facility to complete this project.

REFERENCES

- [1] Cohn, S.B., "Microwave Bandpass Filters Containing High Q Dielectric Resonators," IEEE Transactions on Microwave Theory & Techniques, vol. 16, April 1968, pp. 218-227.
- [2] Fiedziuszko, S.J., "Microwave Dielectric Resonators," Microwave Journal, Sept. 1986, pp. 189-200.
- [3] McAllister, M.W., and S.A. Long, "Rectangular Dielectric Resonator Antenna," IEEE Letters, vol. 19, March 1983, pp. 218-219.
- [4] Mongia, R.K., and A. Ittipiboon, "Theoretical and Experimental Investigations on Rectangular Dielectric Resonator Antennas," IEEE Transactions on Antennas & Propagation, Vol. 45, No. 9, Sept. 1997, pp. 1348-1356.

Motion Detection and Tracking of Video Sequences: A Survey

Neha Kumari

Department of Telecommunication Systems Engineering
Amity Institute of Telecom Technology & Management, Amity University, Uttar Pradesh
Noida, U.P, India, nehabtech08@gmail.com

Abstract: Motion detection, the process which segments moving objects in video streams, is the first critical process of the automatic video surveillance system. This survey paper focuses on key steps in video analysis i.e. Detection of moving objects of interest and tracking of such objects from frame to frame. Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. Discuss the methods of tracking techniques such as point tracking, kernel tracking, Silhouette tracking methods. Various object detection and tracking approaches are compared and analyzed.

Keywords: feature selection, image segmentation, object representation, point tracking.

1. INTRODUCTION

Videos are actually sequences of images, each of which called a frame, displayed in fast enough frequency so that human eyes can percept the continuity of its content. Visual content can be modeled as a hierarchy of abstractions. At the first level are the raw pixels with color or brightness information. Further processing yields features such as edges, corners, lines, curves, and color regions. A higher abstraction layer may combine and interpret these features as objects and their attributes.

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Object tracking is to monitor objects spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem, the problem of matching the target region in successive frames of a sequence of images taken at closely-spaced time intervals great deal of interest in object tracking algorithms.

Object Detection and Tracking Approaches

2. OBJECT REPRESENTATION

In a tracking scenario, an object can be defined as anything that is of interest for further analysis. For instance, boats on the sea, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road, or bubbles in the water are a set of objects that may be important to track in a specific domain. Objects can be represented by their shapes and appearances. In this section, we will first describe the object shape representations commonly employed for tracking and then address the joint shape and appearance representations.

Points. The object is represented by a point, that is, the centroid (Figure 1(a)) or by a set of points (Figure 1(b)). In general, the point representation is suitable for tracking objects that occupy small regions in an image.

Primitive geometric shapes. Object shape is represented by a rectangle, ellipse (Figure 1(c), (d)). Object motion for such representations is usually modeled by translation, affine, or projective (homography) transformation. Though primitive geometric shapes are more suitable for representing simple rigid objects, they are also used for tracking non rigid objects.

Object silhouette and contour. Contour representation defines the boundary of an object (Figure 1(g), (h)). The region inside the contour is called the silhouette of the object (see Figure 1(i)). Silhouette and contour representations are suitable for tracking complex no rigid shapes.

Articulated shape models. Articulated objects are composed of body parts that are held together with joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. The relationship between the parts is governed by kinematic motion models, for example, joint angle, etc. In order to represent an articulated object, one can model the constituent parts using cylinders or ellipses as shown in Figure 1(e).

Skeletal models. Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for

recognizing objects. Skeleton representation can be used to model both articulated and rigid objects (see Figure 1(f)).

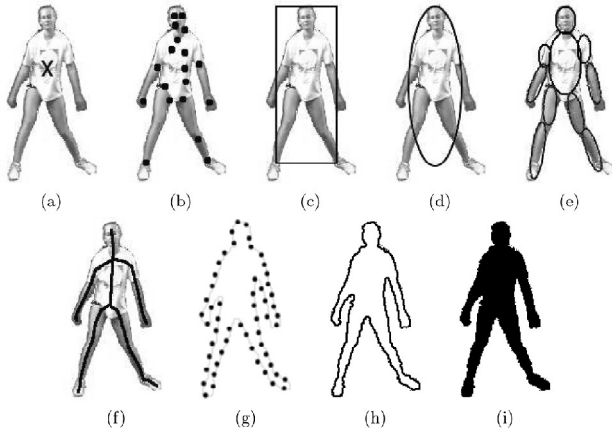


Fig. 1. Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette.

3. FEATURE SELECTION FOR TRACKING

Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness so that the objects can be easily distinguished in the feature space. Feature selection is closely related to the object representation. For example, color is used as a feature for histogram-based appearance representations, while for contour-based representation, object edges are usually used as features. In general, many tracking algorithms use a combination of these features. The details of common visual features are as follows:

Color. The apparent color of an object is influenced primarily by two physical factors, the spectral power distribution of the illuminant and the surface reflectance properties of the object. In image processing, the RGB (red, green, blue) color space is usually used to represent color. However, the RGB space is not a perceptually uniform color space, that is, the differences between the colors in the RGB space do not correspond to the color differences perceived by humans. Additionally, the RGB dimensions are highly correlated. In contrast, $L^*u^*v^*$ and $L^*a^*b^*$ are perceptually uniform color spaces, while HSV (Hue, Saturation, Value) is an approximately uniform color space. However, these color spaces are sensitive to noise. In summary, there is no last word on which color space is more efficient, therefore a variety of color spaces have been used in tracking.

Edges- Object boundaries usually generate strong changes in image intensities. Edge detection is used to identify these changes. An important property of edges is that they are less

sensitive to illumination changes compared to color features. Algorithms that track the boundary of the objects usually use edges as the representative feature. Because of its simplicity and accuracy, the most popular edge detection approach is the Canny Edge detector. An evaluation of the edge detection algorithms is provided by:

Optical Flow is a dense field of displacement vectors which defines the translation of each pixel in a region. It is computed using the brightness constraint, which assumes brightness constancy of corresponding pixels in consecutive frames. Optical flow is commonly used as a feature in motion-based segmentation and tracking applications.

Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. Compared to color, texture requires a processing step to generate the descriptors. There are various texture descriptors:

Gray-Level Co occurrence Matrices (GLCM's) (a 2D histogram which shows the co occurrences of intensities in a specified direction and distance), Law's texture measures (twenty-five 2D filters generated from five 1D filters corresponding to level, edge, spot, wave, and ripple), wavelets (orthogonal bank of filters), and steerable pyramids. Similar to edge features, the texture features are less sensitive to illumination changes compared to color. Mostly features are chosen manually by the user depending on the application domain. However, the problem of automatic feature selection has received significant attention in the pattern recognition community. Automatic feature selection methods can be divided into *filter* methods and *wrapper* methods. The filter methods try to select the features based on a general criteria, for example, the features should be uncorrelated. The wrapper methods select the features based on the usefulness of the features in a specific problem domain, for example, the classification performance using a subset of features.

Among all features, color is one of the most widely used feature for tracking. Despite its popularity, most color bands are sensitive to illumination variation. Hence in scenarios where this effect is inevitable, other features are incorporated to model object appearance. Alternatively, a combination of these features is also utilized to improve the tracking performance

Table I: Object Detection Categories

Categories	Representative Work
Point Detectors	Moravec's detector, Harris detector , Scale Invariant Feature Transform , Affine Invariant Point Detector

Segmentation	Mean-shift, Graph-cut, Active contours.
Background Modeling	Mixture of Gaussians, Eigen background Wall flower Dynamic texture background.
Supervised Classifiers	Support Vector Machines, Neural Networks, Adaptive Boosting.

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracks.

4. POINT DETECTORS

Point detectors are used to find interest points in images which have an expressive texture in their respective localities. Interest points have been long used in the context of motion, stereo, and tracking problems. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint. In the literature, commonly used interest point detectors include Moravec's interest operator, Harris interest point detector, KLT detector, and SIFT detector as illustrated in figure 2.

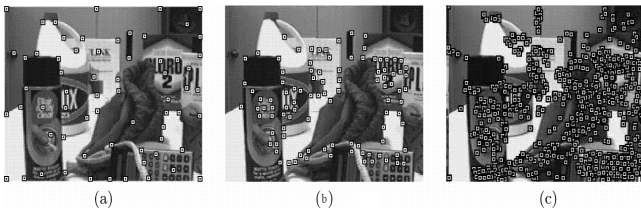


Fig-2 Interest points detected by applying (a) the Harris, (b) the KLT, and (c) SIFT operators

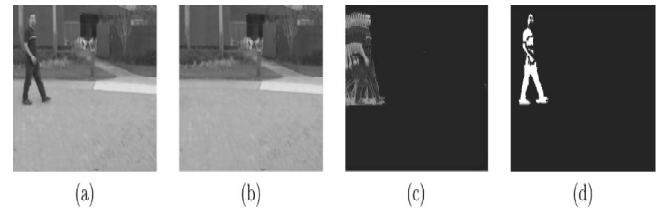
of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks.

Point detectors are used to find interest points in images which have an expressive texture in their respective localities. Interest points have been long used in the context of motion, stereo, and tracking problems. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint. In the literature, commonly used interest point detectors include Moravec's

interest operator, Harris interest point detector, KLT detector, and SIFT detector as illustrated in figure 2.

5. BACKGROUND SUBTRACTION

Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the *background subtraction*. For instance, Stauffer and Grimson use a mixture of Gaussians to model the pixel color. In this method, a pixel in the current frame is checked against the background model by comparing it with every Gaussian in the model until a matching Gaussian is found. If a match is found, the mean and variance of the matched Gaussian is updated, otherwise a new Gaussian with the mean equal to the current pixel color and some initial variance is introduced into the mixture. Each pixel is classified based on whether the matched distribution represents the background process. Moving regions, which are detected using this approach, along with the background models are shown in Figure 3.



Mixture of Gaussian modeling for background subtraction. (a) Image from a sequence in which a person is walking across the scene. (b) The mean of the highest-weighted Gaussians at each pixels position. These means represent the most temporally persistent per-pixel color and hence should represent the stationary background. (c) The means of the Gaussian with the second-highest weight; these means represent colors that are observed less frequently. (d) Background subtraction result. The foreground consists of the pixels in the current frame that matched a low-weighted Gaussian.

Another approach is to incorporate region-based (spatial) scene information instead of only using color-based information. Elgammal and Davis use nonparametric kernel density estimation to model the per-pixel background. During the subtraction process, the current pixel is matched not only to the corresponding pixel in the background model, but also to the nearby pixel locations. Thus, this method can handle camera jitter or small movements in the

background. Li and Leung [2002] fuse the texture and color features to perform background subtraction over blocks of 5×5 pixels. Since texture does not vary greatly with illumination changes, the method is less sensitive to illumination. Toyama et al. [1999] propose a three-tiered algorithm to deal with the background subtraction problem. In addition to the pixel-level subtraction, the authors use the region and the frame-level information. At the pixel level, the authors propose to use Wiener filtering to make probabilistic predictions of the expected background color. At the region level, foreground regions consisting of homogeneous color are filled in. At the frame level, if most of the pixels in a frame exhibit sudden change, it is assumed that the pixel-based color background models are no longer valid. At this point, either a previously stored pixel-based background model is swapped in, or the model is reinitialized. The foreground objects are detected by projecting the current image to the Eigen space and finding the difference between the reconstructed and actual images. We show detected object regions using the Eigen space approach in Figure 4.

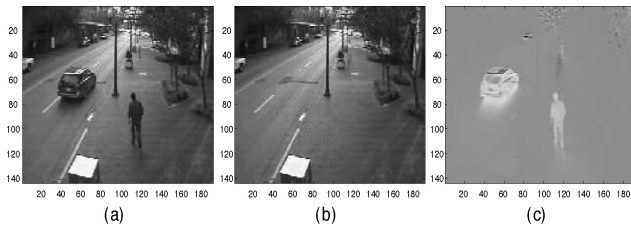


Fig. 4. Eigenspace decomposition-based background subtraction (space is constructed with objects in the FOV of camera): (a) an input image with objects, (b) reconstructed image after projecting input image onto the eigenspace, (c) difference image. Note that the foreground objects are clearly identifiable.

6. SEGMENTATION

The aim of image segmentation algorithms is to partition the image into perceptually similar regions. Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving efficient partitioning.

A. Mean-Shift Clustering. For the image segmentation problem, Comaniciu and Meer [2002] propose the mean-shift approach to find clusters in the joint spatial color space, $[l, u, v, x, y]$, where $[l, u, v]$ represents the color and $[x, y]$ represents the spatial location. Given an image, the algorithm is initialized with a large number of hypothesized cluster centers randomly chosen from the data. Then, each cluster center is moved to the mean of the data lying inside the multidimensional ellipsoid centered on the cluster center. The vector defined by the old and the new cluster centers is called the *mean-shift vector*. The mean-shift vector is computed iteratively until the cluster centers do not change their positions. Note that during the mean-shift iterations,

some clusters may get merged. In Figure 5(b), we show the segmentation using the mean-shift approach generated using the source code available at Mean Shift Segments.

B. Image Segmentation Using Graph-Cuts. Image segmentation can also be formulated as a graph partitioning problem, where the vertices (pixels), $\mathbf{V} = \{u, v, \dots\}$, of a graph (image), \mathbf{G} , are partitioned into N disjoint subgraphs (regions), $A_i, \dots, A_N, i = 1 \dots N, A_i \cap A_j = \emptyset, i \neq j$, by pruning the weighted edges of the graph. The total weight of the pruned edges between two sub graphs is called a *cut*. The weight is typically computed by color, brightness, or texture similarity between the nodes. Wu and Leahy [1993] use the minimum cut criterion, where the goal is to find the partitions that minimize a cut. In their approach, the weights are defined based on the color similarity. One limitation of minimum cut is its bias toward over segmenting the image. This effect is due to the increase in cost of a cut with the number of edges going across the two partitioned segments

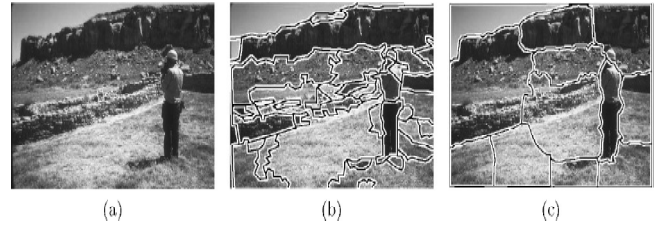
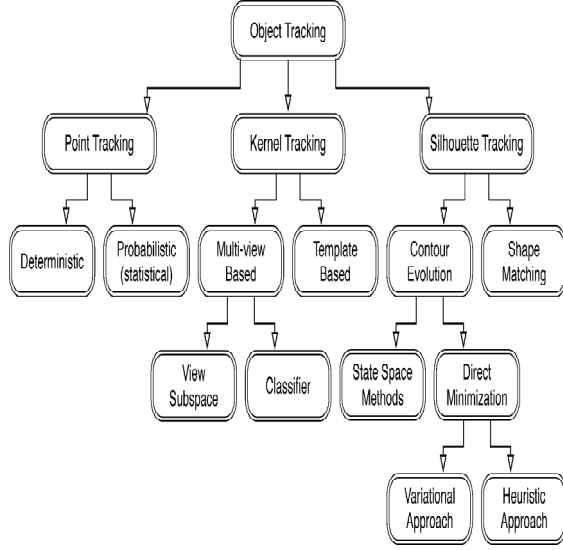


Fig. 5. Segmentation of the image shown in (a), using mean-shift segmentation (b) and normalized cuts (c).

7. OBJECT TRACKING

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. Object tracker may also provide the complete region in the image that is occupied by the object at every time instant. The tasks of detecting the object and establishing correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds objects across frames. In the latter case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames. In either tracking approach, the objects are represented using the shape and/or appearance models described in Section 2. The model selected to represent object shape limits the type of motion or deformation it can undergo. For example, if an object is represented as a point, then only a translational model can be used. In the case where a geometric shape representation like an ellipse is used for the object, parametric motion models like affine or projective transformations are appropriate.



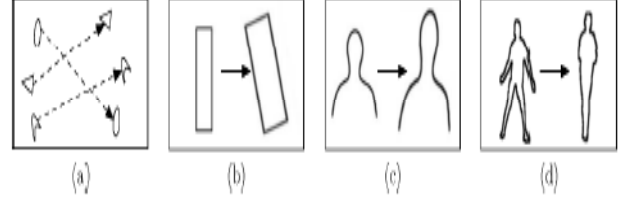
8. TAXONOMY OF TRACKING METHOD

In view of the aforementioned discussion, we provide taxonomy of tracking methods in Figure 6. Representative work for each category is tabulated in Table II. We now briefly introduce the main tracking categories, followed by a detailed section on each category.

Categories	Representative Work
Point Tracking	
Deterministic methods	MGE tracker, GOA tracker.
Statistical methods	Kalman filter, PMHT.
Kernel Tracking	
Template and density based appearance models	Mean-shift, KLT, Layering.
Multi-view appearance models	Eigen tracking [Iack and jepon [1998]; SVM tracker.
Silhouette Tracking	
Contour evolution	State space models, Variational methods, Heuristic methods.
Matching shapes	Hausdorff Histogram.

Point Tracking Objects detected in consecutive frames are represented by points, and the association of the points is based on the previous object state which can include object position and motion. This approach requires an external mechanism to detect the objects in every frame. An example of object correspondence is shown in Figure 7(a).

Fig. 7 (a) Different tracking approaches. Multipoint correspondence, (b) parametric transformation of a rectangular patch, (c, d) Two examples of contour evolution.



A. Kernel Tracking.

Kernel based object tracking using color histogram technique has been applied for different challenging situations. Kernel refers to the object shape and appearance. For example, the kernel can be a rectangular template or an elliptical shape with an associated histogram. Objects are tracked by computing the motion of the kernel in consecutive frames (Figure 7(b)). This motion is usually in the form of a parametric transformation such as translation, rotation, and affine.

B. Silhouette Tracking is performed by estimating the object region in each frame. Silhouette tracking methods use the information encoded inside the object region. This information can be in the form of appearance density and shape models which are usually in the form of edge maps. Given the object models, silhouettes are tracked by either shape matching or contour evolution (see Figure 7(c), (d)). Both of these methods can essentially be considered as object segmentation applied in the temporal domain using the priors generated from the previous frames.

9. CONCLUSION

In this paper, the survey of object detection and tracking methods is presented in this paper. The importance of object shape representations for detection and tracking systems, we have included discussion on popular methods not all the methods for the same the point trackers require detection in every frame, whereas geometric region or contours-based trackers require detection only when the object first appears in the scene. Recognizing the importance of object detection for tracking systems, we include a short discussion on popular object detection methods. A detailed summary of criteria for feature selection, object tracking methods is presented which can give valuable insight into this important research topic.

REFERENCES

- [1] GREGORY D. HAGER and Peter N. Belhumeur ,Efficient Region Tracking With Parametric Models of Geometry and Illumination.IEEE transactions on pattern analysis and machine intelligence, vol. 20, no. 10, pp. 1025-39 October 1998.
- [2] VEENMAN, C., REINDERS, M., AND BACKER, E. 2001. Resolving motion correspondence for densely moving points. IEEE Trans. Patt. Anal. Mach. Intell. 23, 1, 54–72.

- [3] SERBY, D., KOLLER-MEIER, S., AND GOOL, L. V. 2004. Probabilistic object tracking using multiple features. In *IEEE International Conference of Pattern Recognition (ICPR)*. 184–187.
- [4] CHUNHUA SHEN, JUNAE KIM AND HANZI WANG, Generalised Kernel –Based visual Tracking .*IEEE transaction on circuit and system for video technology*, Vol.20,no.1,January 2010
- [5] YILMAZ, A., LI, X., AND SHAH, M. 2004. Contour based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Trans. Patt. Analy. Mach. Intell.* 26, 11, 1531–1536.
- [6] BALLARD, D. AND BROWN, C. 1982. *Computer Vision*. Prentice-Hall.
- [7] ALI, A. AND AGGARWAL, J. 2001. Segmentation and recognition of continuous human activity. In *IEEE Workshop on Detection and Recognition of Events in Video*. 28–35.
- [8] PASCHOS, G. 2001. Perceptually uniform color spaces for color texture analysis: an empirical evaluation. [19] *IEEE Trans. Image Process.* 10, 932–937.
- [9] SONG, K. Y., KITTLER, J., AND PETROU, M. 1996. Defect detection in random color textures. *Israel Verj. Cap.J.* 14, 9, 667–683
- [10] CANNY, J. 1986. A computational approach to edge detection. *IEEE Trans. Patt. Analy. Mach. Intell.* 8, 6, 679–698
- [11] BOWYER, K., KRANENBURG, C., AND DOUGHERTY, S. 2001. Edge detector evaluation using empirical roc curve. *Comput. Vision Image Understand.* 10, 77–103
- [12] SYNH VIET-UYEN HA AND JAE WOOK JEON. Readjusting Unstable Regions to Improve the Quality of High Accuracy Optical Flow. *IEEE transaction on circuit and system for video technology*, Vol. 20, NO. 4, APRIL 2010
- [13] HARALICK, R., SHANMUGAM, B., AND DINSTEIN, I. 1973. Textural features for image classification. *IEEE Trans.Syst. Man Cybern.* 33, 3, 610–622
- [14] LAWS, K. 1980. Textured image segmentation. PhD thesis, Electrical Engineering, University of Southern California
- [15] MALLAT, S. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Analy. Mach. Intell.* 11, 7, 674–693.
- [16] GREENSPAN, H., BELONGIE, S., GOODMAN, R., PERONA, P., RAKSHIT, S., AND ANDERSON, C. 1994. Overcomplete steerable pyramid filters and rotation invariance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 222–228.
- [17] BLUM, A. L. AND LANGLEY, P. 1997. Selection of relevant features and examples in machine learning. *Artific.Intell.* 97, 1-2, 245–271
- [18] MORAVEC, H. 1979. Visual mapping by a robot rover. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 598–600
- [19] HARRIS, C. AND STEPHENS, M. 1988. A combined corner and edge detector. In *4th Alvey Vision Conference*. 147–151.
- [20] SHI, J. AND TOMASI, C. 1994. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 593–600.
- [21] LOWE, D. 2004. Distinctive image features from scale-invariant key points. *Int. J. Comput. Vision* 60, 2, 91–110.
- [22] ELGAMMAL, A., HARWOOD, D., AND DAVIS, L. 2000. Non-parametric model for background subtraction. In *European Conference on Computer Vision (ECCV)*. 751–767.
- [23] SHI, J. AND MALIK, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Patt. Analy. Mach. Intell.* 22, 8, 888–905.
- [24] MORAVEC, H. 1979. Visual mapping by a robot rover. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 598–600
- [25] CASELLES, V., KIMMEL, R., AND SAPIRO, G. 1995. Geodesic active contours. In *IEEE International Conference on Computer Vision (ICCV)*. 694–699.
- [26] STAUFFER, C. AND GRIMSON, W. 2000. Learning patterns of activity using real time tracking. *IEEE Trans. Patt. Analy. Mach. Intell.* 22, 8, 747–767.
- [27] COMANICIU, D. AND MEER, P. 1999. Mean shift analysis and applications. In *IEEE International Conference on Computer Vision (ICCV)*. Vol. 2. 1197–1203.
- [28] OLIVER, N., ROSARIO, B., AND PENTLAND, A. 2000. A bayesian computer vision system for modeling human interactions. *IEEE Trans. Patt. Analy. Mach. Intell.* 22, 8, 831–843
- [29] TOYAMA, K., J. KRUMM, B. B., AND MEYERS, B. 1999. Wallflower: Principles and practices of background maintenance. In *IEEE International Conference on Computer Vision (ICCV)*. 255–261.
- [30] MONNET, A., MITTAL, A., PARAGIOS, N., AND RAMESH, V. 2003. Background modeling and subtraction of dynamic scenes. In *IEEE International Conference on Computer Vision (ICCV)*. 1305–1312.
- [31] PAPAGEORGIOU, C., OREN, M., AND POGGIO, T. 1998. A general framework for object detection. In *IEEE International Conference on Computer Vision (ICCV)*. 555–562.
- [32] ROWLEY, H., BALUJA, S., AND KANADE, T. 1998. Neural network-based face detection. *IEEE Trans. Patt. Analy. Mach. Intell.* 20, 1, 23–38.
- [33] VIOLA, P., JONES, M., AND SNOW, D. 2003. Detecting pedestrians using patterns of motion and appearance. In *IEEE International Conference on Computer Vision (ICCV)*. 734–741.
- [34] SALARI, V. AND SETHI, I. K. 1990. Feature point correspondence in the presence of occlusion. *IEEE Trans. Patt. Analy. Mach. Intell.* 12, 1, 87–91.
- [35] BROIDA, T. AND CHELLAPPA, R. 1986. Estimation of object motion parameters from noisy images. *IEEE Trans. Patt. Analy. Mach. Intell.* 8, 1, 90–99.
- [36] STREIT, R. L. AND LUGINBUHL, T. E. 1994. Maximum likelihood method for probabilistic multi-hypothesis tracking. In *Proceedings of the International Society for Optical Engineering (SPIE.)* vol. 2235. 394–405.
- [37] TAO, H., SAWHNEY, H., AND KUMAR, R. 2002. Object tracking with bayesian estimation of dynamic layer representations. *IEEE Trans. Patt. Analy. Mach. Intell.* 24, 1, 75–89.

13T Low Power PTL based Arithmetic Leaf Cell for Signal Processing

S.Vijayakumar¹, Reeba Korah²

¹Dept. of Electronics and Communication Engineering
Ganadipathy Tulsi's Jain Engineering College, Vellore, India – 632102.
vijaysuresh1975@yahoo.com

²Dept. of Electronics and Communication Engineering
St. Joseph's College of Engineering, Chennai, India – 600119.
reeba26in@gmail.com

Abstract: Though the technology hunt is on for sub - 10nm, the circuit level abstraction is still proving as the eminent way to reduce power in any form of digital design. In this work, one bit full adder cell is implemented using the pass transistor method as an arithmetic leaf cell. The CCMOS, BBL-PT, ULPFA, SERF and Transmission Gate Logic styles are taken for comparison. The results show that the proposed design operates at low power than the other methods which are compared. The power consumed by the new 13T adder as a leaf cell is 33 times less as the maximum optimization when compared to BBL-PT adder.

Index Terms: 13T adder, low power, PTL adder, 1bit CCMOS adder, power efficient, arithmetic leaf cell.

1. INTRODUCTION

Minimizing power, delay and area are the important constraints in all kind of digital VLSI design. But all the factors are impossible to meet in parallel. The growing global population demands the need for productivity of gadgets which are to be handy and cheaper. The device scaling is the key technique for miniaturization of such digital equipments.

There are many advantages of scaling for low-power operation. The improved device characteristic for low-voltage operation is one of them. This is due to the improvement of the current driving capabilities. Improved interconnect technology is used to minimize the parasitic effects. Reduced junction capacitance is another result of scaling the device. Availability of multiple and variable threshold devices lead to MTCMOS technology. This results in good management of active and standby power trade-off and higher density of integration. The sleep transistor and power gating are the familiar methods in connection with scaling the devices.

However, as the size of CMOS circuits become smaller, high performance is more difficult to achieve due to the effect of velocity saturation [1] – [2]. At short channel lengths L , the

drain current per channel width (W) is no longer proportional to $(V_{gs} - V_t)^2/L$, where V_{gs} is the gate-to-source voltage and V_t is the threshold voltage. The drain current is proportional to $V_{sat}(V_{gs} - V_t)$, where V_{sat} is the saturation velocity [3] – [4]. Moreover, at sub-micron level the difference between the operating voltage and threshold voltage is very small and the static leakage is dominant to increase the static power consumption. This situation increases the design complexity and careful design of the system becomes more dominant factor. However the technology scaling has already reached the saturation level. The current scenario looks ahead the alternate solution to the VLSI industry to keep the Moore's law to be true. Power optimization with a satisfied delay and area or vice-versa can be achieved with the help of various abstraction levels. Out of such abstraction levels, the circuit level approach is the competent one to reduce the power despite thinking the technology scaling alone. Though the circuit approach seems to be very complicated and more time consuming to design or re-engineer the desired modules, the through-put is good at the final stage.

In general, the power dissipation in a CMOS circuit is expected to be smaller along with a minimum delay. Technology, circuit design style, architectural abstraction and algorithmic abstraction are the factors which influence the power dissipation in CMOS circuit. CMOS circuit is robust as it is proven already due to its pull-up and pull-down networks which ensure the full swing and hence immune to noise factors. Numerous CMOS circuit design styles exist which are static and dynamic in nature. The following paragraphs brief the logic styles available to construct a 1-bit adder cell.

The full adder employed with PMOS transistors to allow the output to be charged high and NMOS transistors to discharge the output node to ground, is of CMOS style. The PMOS pull-up networks are hence used to ensure the output to a full logic high level. The NMOS pull-down networks, on the other end are used to ensure the output to a full logic low level. All together, the CMOS circuit operates to

produce a full swing output voltage. NO RAcE dynamic CMOS (NORA) is another method of constructing a full adder. The P-type stage which forms the carry output is dynamically pre-charged high while the N-type tree which computes the sum is pre-discharged to ground. This process requires a two phase clock. Cascode voltage switch logic (CSVL) is another method to design a full adder. It is a dynamic logic family like a NORA. But it never requires a complementary clock.

Another alternative method to the CMOS static complementary logic is the conventional pass-transistor logic based on MOS switches. It consists of a complementary pair connected in parallel. It acts as a switch, with the logic variable A as the control input. If A is low, the gate is OFF and presents high resistance between the terminals. If A is high, the gate is ON and acts as a switch with an on resistance of R_{on} . The literature review on a 1-bit full adder cell is given in section II along with the other adder methods which are considered for this work. The implementation of the proposed – new 13T adder is to be given in section III. The simulation and performance comparison in terms of average power and delay are available in section IV. The work is concluded in section V.

2. PREVIOUS 1-BIT FULL ADDERS

In general, a one bit full adder has three 1-bit inputs A , B , and C_{in} which are required to calculate the two 1-bit outputs Sum and C_{out} . They are expressed as,

$$Sum = (A \oplus B) \oplus C_{in} \quad (1)$$

$$C_{out} = A.B + C_{in}(A \oplus B) \quad (2)$$

The selection of 1 bit adders for the comparison is based on the performance metrics of the methods which depend on power, delay and area. Another important factor is to consider the full swing output which yields a robust adder design. The module operates with lower switching activity consumes low power. The circuit with less number of nodes has the possibility of faster propagation. Among all such considerations, the design with less device count is also taken into account which leads to less chip area. Hence CCMOS, TFA, TGA, BBL-PT, ULPFA and SERF adder are taken for analyzing along with the proposed design due to the competency of these logic styles in terms of aforesaid factors.

A. Regular CMOS Adder

The Complementary static CMOS (CCMOS) adder is termed as regular adder which uses the PMOS pull-up and NMOS pull-down networks (complement to each other) with a minimum possible MOS transistors as in Fig. 1. The

expressions to construct sum and carry are derived from (1), (2) as

$$S = AB'C_{in}' + A'BC_{in}' + A'B'C_{in} + ABC_{in} \quad (3)$$

$$Co = AB + BC_{in} + AC_{in} \quad (4)$$

Where S is the Sum output and Co is the Carry output and the inputs with a single punctuation (') are the complementary inputs. To reduce the device count, (3) is simplified as

$$S = AB'C_{in} + C_o'(A+B+C_{in}) \quad (5)$$

Equation (5) is required to reduce the transistor count with the help of logic sharing between the sum and carry generation circuits [5]-[6]. This is also in other words useful to reduce power consumption due to reduced device count. Substituting (4) into (5),

$$S = AB'C_{in} + (AB + BC_{in} + AC_{in})'(A+B+C_{in}) \quad (6)$$

$$= AB'C_{in} + (AB)'(BC_{in})'(AC_{in})'(A+B+C_{in})$$

$$= AB'C_{in} + (A'+B')(B'+C_{in}')(A'+C_{in}')(A+B+C_{in})$$

Solving the expressions results in (3). Hence the logic equivalence can be easily verified. Though the logic sharing is used in the design, it still requires large area with 28 transistors. In the carry generation circuit, the signal propagates through two inverting stages. This leads to an increased delay in the carry path. Moreover it is an important factor to be carefully handled to keep it in control. The reason is that in an arithmetic circuit, the carry rippling to higher stages of a multiplier is the critical path [7].

B. Transmission Function Adder

The next method is the Transmission Function full Adder (TFA) which is as in Fig. 2. It consists of 16 transistors. The basic XOR circuit used in the design requires one of its two inputs in complementary forms. An additional inverter is thus needed, which leads to a 6-transistor XOR design. The design employs 4 CMOS transmission gates and can achieve full voltage swing operations and also operates at low power [8]-[10]. The main disadvantage of this logic styles is that it lacks the driving capability. When TFA is used in cascaded stages, its performance degrades significantly due to threshold voltage decay.

C. Transmission Gate Adder

The third logic design is denoted as TG-CMOS. It contains 20 transistors and uses only transmission gates and inverters to implement XOR and multiplexing functions. It is a widely used solution to deal with the voltage drop problem. NMOS

transistor passes a strong 0 but a weak 1 and the PMOS does a strong 1 and pass a weak 0. CCMOS uses a PMOS pull up and an NMOS pull down whereas the transmission gate combines the best of both device advantages [11] – [14] by placing an NMOS device in parallel with a PMOS as in Fig. 3.

The transmission gate acts as a bi-directional switch controlled by gate signal C as shown in Fig. 4(a). When $C = 1$, both inputs are driven, allowing the signal to pass through the gate. It is expressed as,

$$A = B \text{ for } C = 1 \quad (7)$$

Otherwise the PMOS and NMOS are at cut-off state.

To understand the function of a transmission gate, consider the charging node B to V_{dd} as in Fig. 4(b). Node A is set to V_{dd} and the transmission gate is enabled ($C = 1$ and $C' = 0$). Since both the NMOS & PMOS transistors are on, the node B is fully charged to V_{dd} . On the other hand, while discharging the node B to 0, B is initially at V_{dd} when the node A is driven low. The PMOS can pull-down the node B to V_{TP} at which moment it turns off. The parallel NMOS device stays turned on and at $V_{GS} = V_{dd}$ the node B is pulled down to GND . Though the transmission gate needs two transistors to achieve this condition, it ensures a full swing.

D. BBL – PT Adder

This is a hybrid full adder which has the combination of Branch Based Logic and Pass Transistor circuits (BBL – PT). Low power consumption is the prime objective of this method. The circuit is implemented with a few transistors and less intra-cell node connections as possible. Hence the structure is implemented using branch-based design technique which meets the requirements while ensuring robustness with respect to voltage and device scaling.

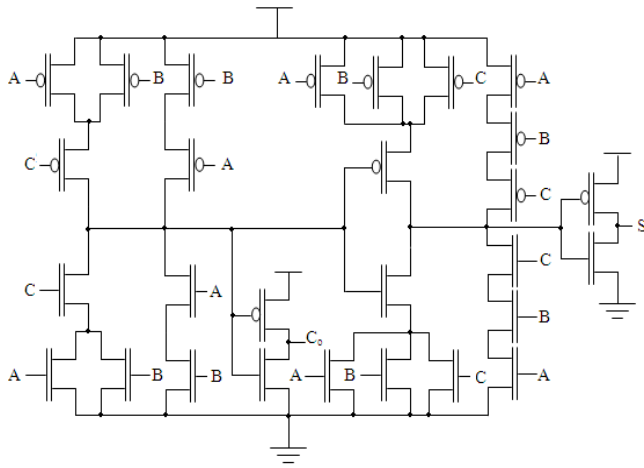


Fig. 1. Regular CMOS full adder

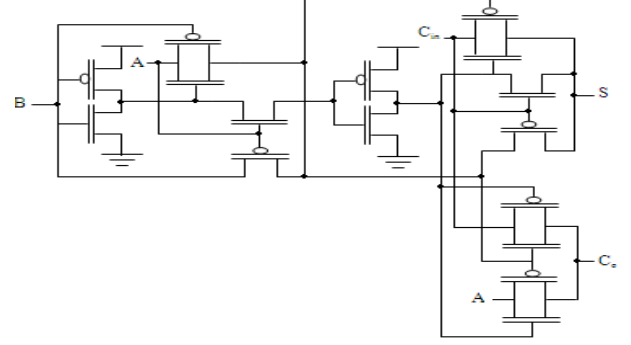


Fig. 2. TFA full adder

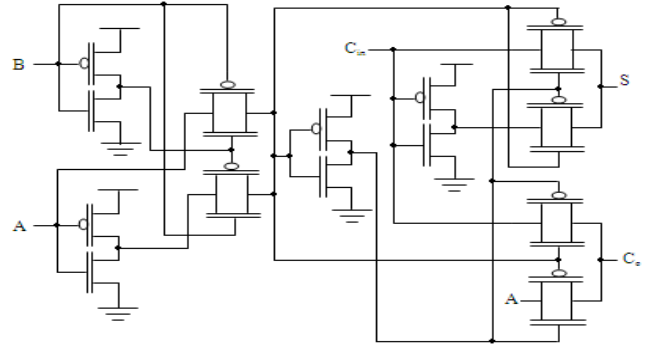


Fig. 3. TGA full adder

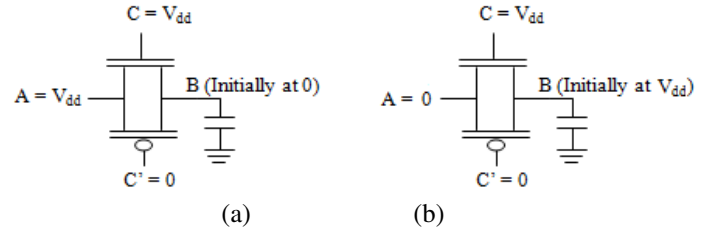


Fig. 4. Functions of transmission gate: (a) Charging node B, (b) Discharging node B.

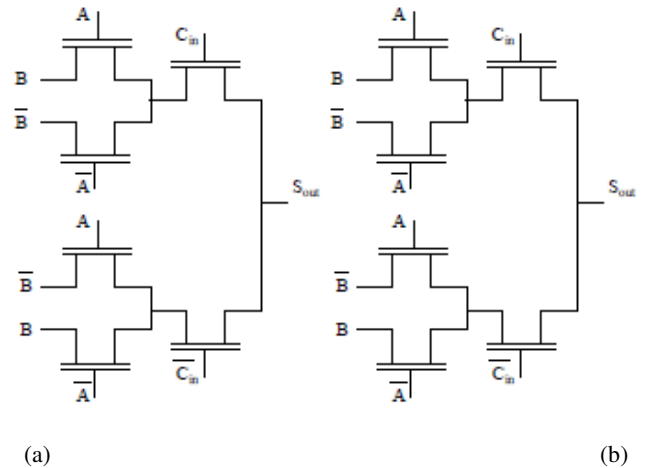


Fig. 5. BBL – PT full adder: (a) Sum circuit, (b) Carry circuit.

At the physical design level, this branch-based design diminishes the diffusion capacitance since it eases diffusion-sharing which leads to regular and compact layout. Series connection of transistors between the output node and the supply rail forms the branch like circuit and hence the name BBL. Using Karnaugh's maps, the circuit is optimized with PMOS and NMOS networks to produce the sum of products of the Boolean expression.

The carry block is constructed using branch based logic as in fig. 5(b). But the sum block doesn't follow this method of implementation, because it requires 24 transistors to construct. Pass transistor logic (PT) is used to achieve the functionality with less device count [15]. The sum block with the PT structure is hence computing the arithmetic function $AB + BC_{in} + AC_{in}$ which is shown in fig. 5(a). The one bit adder of this type is therefore termed as BBL-PT, because it uses the BBL method to compute carry and PT method to estimate the sum output.

E. ULPFA

It is an Ultra-Low Power Full Adder (ULPFA) which consists of Low Power (LP) XOR gate. The sum logic is achieved by connecting the PMOS and NMOS transistors in a way similar to complementary pass transistor style. Two PMOS transistors in series and 2 NMOS transistors arranged in a manner along with an inverter forms a two input XOR gate. This arrangement is again replicated and connected with it to form the circuit of a three input XOR gate as shown in fig. 6. The circuit to compute the carry output signal is similar to the carry computation circuit given for a BBL-PT adder as in fig. 5(b).

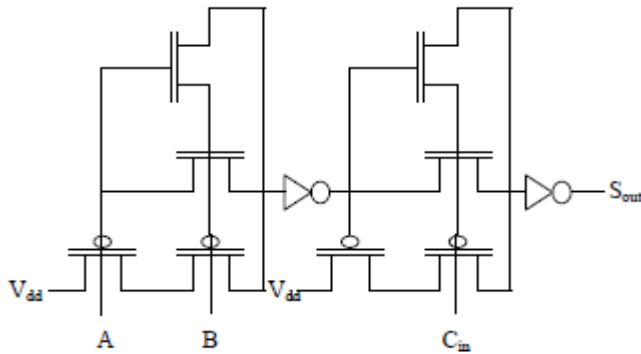


Fig. 6. Sum Circuit of ULPFA

This method operates at a supply voltage less than 1 volt and hence the design is named as Ultra-Low Power Full Adder. It operates consistently even under the voltage less than 0.6V as mentioned in [15]. Another benefit of this logic style is that the leakage power is negligible than the consumption by a MOSFET. But the delay of the sum circuit worsens at the supply voltage less than 0.8V.

F. SERF

The other logic style which competitively consumes lower power is the Static Energy Recovery Full adder (SERF). In the Conventional CMOS, the load capacitance charged to Vdd is discharging to ground when the output is switching from logic '1' to logic '0'. This kind of non-energy recovery leads to higher short circuit power P_{sc} . Whereas in an energy recovery circuit, the charge stored at the load capacitor (C_L) is recycled to the circuit and hence the energy recovery full adder is a power efficient design as represented in fig. 7. This idea leads to a full adder with less transistor count and tremendous decrease in total power utilization.

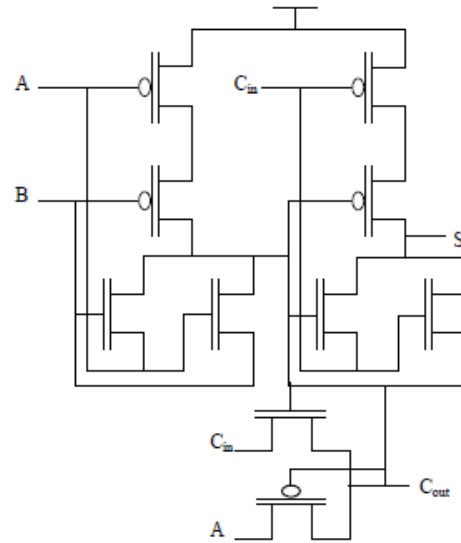


Fig. 7. Static Energy Recovery Full adder (SERF)

The SERF adder design needs only 10 transistors to construct the complete 1-bit adder logic unit which is the lowest among all the other styles described so far. Hence it leads to the requirement of lowest die area. The transient function of this kind of adder is faster and competent with the other logic methods taken for analysis. But it has the problem of threshold loss, i.e., the circuit is producing the degraded output. The use of this type of adder hence becomes critical when it is used to build larger system.

3. NEW 13T – PTL ADDER

A. Pass Transistor Logic (PTL)

Conventionally the inputs are applied to gate terminals of transistors. In pass-transistor circuits, the inputs are also applied to source and or drain terminals. The Pass Transistor Logic is generally constructed with the help of NMOS alone or parallel combination of PMOS and NMOS as transmission gates. The transistors in such circuits are working as switches and are providing the control over the parts of entire circuit. Implementation of complex gates can

be possible with minimum number of transistors which results in a simpler circuit and the node capacitance is also reduced [16].

Pass transistor is a ratio-less logic class. The dc characteristics are not affected by the size of transistor. Increasing the size reduces the resistance but this can be offset by the increase in diffusion capacitance. Pass transistors in a cascaded chain can be arranged from largest to smallest to reduce delay. The pass transistors have a series resistance associated with them.

B. Proposed 13T PTL design

Yet again there are enough chances available to optimize a digital VLSI design with circuit level abstraction. It is very hard to construct and re-engineer a design at this level. However, doing so carefully will surely yield new dimension to a circuit with optimized performance. The use of reduced logic gates is possible in a circuit if there is an existence of logic sharing. Concurrently, there is a possibility of circuit construction with few transistors using the PTL method.

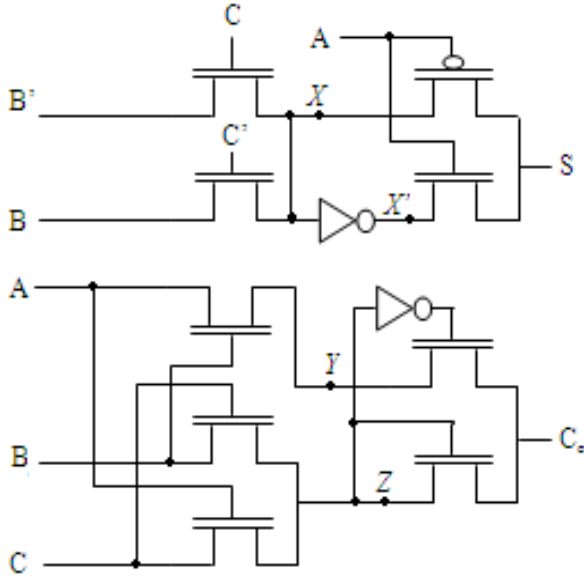


Fig. 8. Proposed 13T - PTL Full Adder

The proposed method is constructed using the simple Boolean based full adder of PTL style. The following few paragraphs have the implementation strategy into two groups namely the sum circuit and the carry computation. In the Sum circuitry, input signals B and C are used to compute a two bit XOR function in the first half with merely two transistors. The node X is used to express it as,

$$X = B'C + BC' \quad (8)$$

On the other hand the node X' is offering an XNOR function which is produced with the help of an inverter at node X. This node is also used to restore the swing with the help of a static CMOS inverter. From (8), the XOR is inverted using CMOS inverter to get XNOR output expressed as,

$$X' = (B'C + BC')' \quad (9)$$

The sum output is the combination of an XOR and an XNOR with the conditions A=0 and A=1 respectively. Hence the output stage of the sum has two transistors with a total of six transistors which is expressed as

$$S = A'X + AX' \quad (10)$$

Using (8) and (9),

$$S = A'(B'C + BC') + A((B'C + BC')') \quad (11)$$

This is also expressed as,

$$S = A \oplus B \oplus C \quad (12)$$

To generate the carry, 7 transistors are needed. The first part of the carry has three transistors to give the functions of AND, OR logic gates. The next part is the output node which results in a carry output of 13T 1-bit full adder.

$$\begin{aligned} \text{At the node Y,} \\ Y &= A.B \end{aligned} \quad (13)$$

$$\begin{aligned} \text{At the node Z,} \\ Z &= B.C + A.C \end{aligned} \quad (14)$$

At the Carry Output, the top most transistor yields $Y.Z'$.

$$\begin{aligned} \text{The carry is expressed as,} \\ C_o &= YZ' + Z \\ &= A.B(B.C + A.C)' + Z \\ &= A.B(C(A+B))' + Z \\ &= A.B(C' + (A'.B')) + Z \\ &= A.B.C' + A.B.(A'.B') + Z \\ &= C'(A.B) + 0 + Z \end{aligned} \quad (15)$$

Substituting Z from (14),

$$C_o = C'A.B + BC + AC \quad (16)$$

Simplifying,

$$C_o = A.B + B.C + A.C \quad (17)$$

All the above two output generations of sum and carry require a new and simple PTL design with only 13 transistors as in fig.8. The outputs in terms of power, delay and PDP have been given in the next section.

4. EXPERIMENTAL RESULTS

The simulation is carried out with the use of Low Power – 90nm Berkeley Predictive Technology model for analyzing the discussed 1 bit adder cells. The circuits are simulated with a supply voltage of 1V. It is because of the performances of the various methods at this potential level are good. All the adders are simulated and performance measures are tabulated for the same test conditions. The length and width of PMOS and NMOS of the adders are kept uniformly like an unit sized inverter for this purpose. The following paragraphs describe and compare the average

power consumption, propagation delay and power delay product of the methods being considered for comparison with the proposed 13T adder.

In the table I, the average power consumption, Propagation delay and Power Delay Product (PDP) are shown for the seven methods. The average power is calculated for the whole adder of each method. But there are two propagation delay measurements per logic style which are tabulated. It is due to the difference in delay between sum logic circuit and carry path in each design style. Hence PDP also has two different values per style.

Table I: Average Power, Delay and PDP

(Vdd=1v)	Adder Type						
	CCMOS	TFA	TGA	BBL-PT	ULPFA	SERF	13T
Average Power (μ W)	1.56	25.51	23.37	48.05	32.92	31.73	1.43
Sum Delay (ns)	4.9031	4.9303	4.9333	3.6222	4.9313	4.9892	4.9954
Carry Delay (ns)	6.2708	4.9998	4.9998	4.9091	5.0389	5.0064	9.9866
PDP - Sum (fWS)	7.65	125.77	115.29	174.04	162.34	158.31	7.14
PDP - Carry (fWS)	9.79	127.54	116.82	235.88	165.88	158.85	14.28

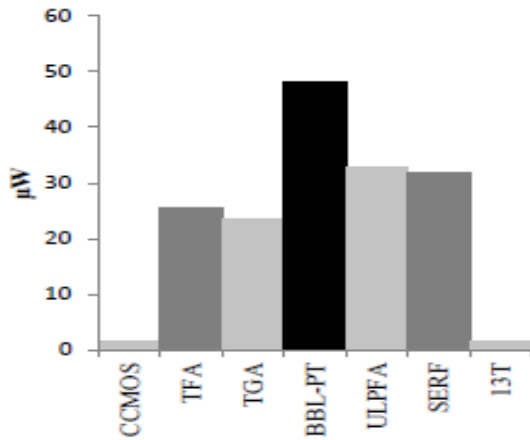


Fig.9. Average power of the various adders

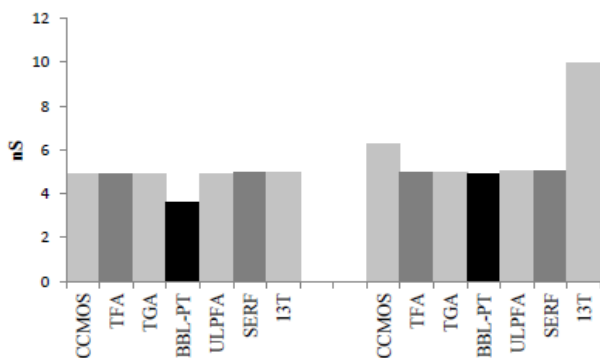


Fig.10. Propagation Delay: (a) Sum-Delay (b) Carry-Delay

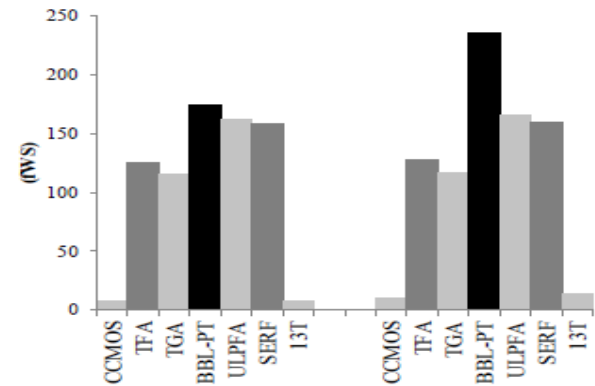


Fig.11. Power Delay Product (PDP: (a) Sum-PDP (b) Carry-PDP

Table – II: Comparison of 13T adder with other Adder styles in percentage

Parameter	Comparison with 13T Leaf Cell in percentage (%)					
	CCMOS	TFA	TGA	BBL-PT	ULPFA	SERF
Average Power	9 ↑	18 X ↑	16 X ↑	33 X ↑	23 X ↑	22 X ↑
Sum-Delay	2 ↓	1 ↓	1 ↓	27 ↓	1 ↓	0.1 ↓
Carry-Delay	37 ↓	50 ↓	50 ↓	51 ↓	49 ↓	50 ↓
Sum-PDP	7 ↑	18 X ↑	16 X ↑	24 X ↑	22 X ↑	22 X ↑
Carry-PDP	31 ↓	9 X ↑	8 X ↑	16 X ↑	11 X ↑	11 X ↑
Device Count	28T ↑	16T ↑	20T ↑	18T ↑	24T ↑	10T ↑

Legends: ↑ - More than 13T, ↓ - Less than 13T, X - Multiple times

The average power consumed by the new 13T - PTL adder is less than the utilization of power by the other methods taken for comparison as in fig.9. The transistor count for CCMOS, TFA, TGA, BBL-PT, ULPFA and SERF are 28T, 16T, 20T, 18T, 24T and 10T respectively which influence the higher power consumption in their respective circuits than the 13T adder. Another advantage of the PTL adder is that the design occupies a less chip area.

Next important factor is the speed of operation of all the methods which are being considered for comparison. To achieve this, the propagation delay should be smaller to compute the output with respect to the applied input. The delay taken by the sum circuit to produce a change in the output is smaller for BBL-PT adder than the other methods as in fig.10(a). The reason is because of the minimum nodes, the signal propagates from its input to output quicker than the rest of the methods. The delay of the carry circuit for the BBL-PT is again less than 13T adder due to the static CMOS used as carry block and hence it is faster as shown in the graph of fig.10(b). The Power delay products for the sum and carry circuits are hence have the relevant values as given in fig.11(a) & fig.11(b) respectively.

From the data shown in the table II, the delay is larger for the proposed adder circuit than the other six methods which is applicable for both sum and carry circuit. The reason is the slower mobility of electrons in a pass transistor circuit than the circuit arrangements of the other logic methods taken for comparison. The following few paragraphs give the reason for the increase in the delay of the PTL adder.

The average velocity of electron is equal to the product of mobility of charge carriers and electric field which is expressed as,

$$v = \mu E \quad (18)$$

Here, E is the electric field defined as

$$E = V_{ds}/L \quad (19)$$

Where, V_{ds} – Drain –Source Voltage,
 L – Channel Length of MOS transistors.

The drain current is expressed as,

$$I_{ds} = Q_{channel} v/L \quad (20)$$

Where, $Q_{channel}$ – Charge across the channel

From equations (18) to (20), it is clear that the velocity of electrons depends on the Electric field which is equal to the drain source voltage [17] – [19]. For the case of 13T, there is signal degradation at the output due to the circuit's dependency on input signal which is not driven by the supply voltage. Hence the potential available in a 13T adder is insufficient to switch the circuit faster like a BBL-PT circuit.

5. CONCLUSIONS

The PTL method of 1 bit full adder cell is implemented in this work along with the CCMOS, TFA TGA, BBL-PT, ULPFA and SERF logic styles to evaluate this type of adder. The objective is met as it is to consume low power which is 33 times less than the BBL-PT adder at the maximum. But the delay of the proposed design is 50% larger than the other methods. This is due to the so called mobility degradation. The other reason is the trade- off between power, delay and area as stated in numerous literatures. The new adder works fine with a minimum PDP in the sum circuit and lack in carry path due to the threshold drop and reduced current driving capabilities. However the PTL adder has the dual advantage of low power and less area. The work may be extended in multipliers to see the scope of its usage in array architectures.

REFERENCES

- [1] J.M.Rabaey, A.Chandrakasan, B.Nikolic, “*Digital Integrated Circuits – A Design Perspective*,” Pearson Educational publishers – 2nd edition 2008.
- [2] N.Weste, and K.Eshraghian, “*Principles of CMOS VLSI design, a System perspective*,” 2nd edition, Addison-Wiley, 1993.
- [3] Y. Taur and T. H. Ning, “*Fundamentals of Modern VLSI Devices*,” Cambridge University Press, Reprint-2004.
- [4] Behzad Razavi, “*Design of Analog CMOS Integrated Circuits*,” Tata McGraw-Hill, 2002.
- [5] Vahid Foroutan, Keivan Navi and Majid Haghighparast, “A New Low Power Dynamic Full Adder Cell Based on Majority Function,” *World Applied Sciences Journal*, 4(1), pp.133-141, 2008.
- [6] Keivan Navi, Omid Kavehie, Mahnoush Rouholamini, Amir Sahafi and Shima Mehrabi, “A Novel CMOS Full Adder,” *20th International Conf. on VLSI Design, VLSID’2007*.
- [7] Keshab K.Parhi, “*VLSI Digital Signal Processing Systems design and Implementation*,” John Wiley & Sons, 1999.
- [8] T.Vigneswaran, B. Mukundhan, and P. Subbarami Reddy, “A Novel Low Power, High Speed 14 Transistor CMOS Full Adder Cell with 50% Improvement in Threshold Loss Problem,” *13th conf. World Academy of Science, Engineering and Technology*, pp.81-85, 2006.
- [9] Nan Zhuang and Haomin Wu, “A New Design of the CMOS Full Adder,” *IEEE Journal of Solid State Circuits*, Vol.27, No.5, pp.840-844, 1992.
- [10] A.M.Shams, Tarek K.Darwish, Magdy A.Bayoumi, “Performance Analysis of low power 1-bit CMOS full adder cells,” *IEEE Tran. On VLSI Systems*, Vol.10 No.1 pp.20-29, Feb’2002.
- [11] Zine Abid, Wei Wang, “New Designs of Redundant binary Full Adders and It’s Applications,” *IEEE Conf. Proc.* pp.3366-3369, 2008.
- [12] Chien-Hung Lin and Shu-Chung Yi, Jin-Jia Chen, “Low Power Adders Design for Portable Video Terminals,” *IEEE Int. Conf. on Intelligent Information Hiding and Signal Processing*, pp.651-654, 2008.

- [13] A.Shams and M.Bayoumi, "Performance Evaluation of 1-bit CMOS adder cells," *Proc. Of IEEE Int. Symposium on Circuit and Systems*, pp.27-30, 1999.
- [14] Kuo-Hsing Cheng, Chih-Sheng hang, "The Novel Efficient Design of XOR/XNOR Function for Adder Applications," *6th IEEE Conf. Proceedings of ICECS '99*, pp. 29-32, 1999.
- [15] Ilham Hassoune, Denis Flandre, Ian O'Connor, Jean-Didier Legat, "ULPFA: A New Efficient Design of a Power-Aware Full Adder," *IEEE Transactions On Circuits And Systems—I*, Vol. 57, No. 8, August 2010.
- [16] A.Chandrakasan, R.W.Brodersen, "Low power Digital CMOS Design," *Kluwer Academic Publishers*, 2002.
- [17] J.M.Rabaey, M.Pedram, "Low Power Design Methodologies," *Kluwer Academic Publishers*, 1996.
- [18] Kaushik Roy, Sharad C.Prasad, "Low Power CMOS VLSI Circuit Design," *John Wiley & Sons*, 2000.
- [19] G.K.Yeap, F.N.Najim, "Low Power VLSI Design and Technology," *World Scientific Publishers*, 1996.

Large Scale Path Loss Outdoor Propagation Models: A Survey

Richa Budhiraja

Dept. of Telecommunication Systems Engineering, Amity Institute of Telecom Technology & Management
Amity University, Uttar Pradesh, Noida, U.P, India, richa_1407@yahoo.co.in

Abstract: Large scale path loss modeling plays a fundamental role in designing both fixed & mobile radio systems. Radio propagation is essential for emerging technologies with appropriate design, deployment and management strategies for any wireless network. It varies significantly depending upon the terrain, channel environment, frequency band and desired radio coverage path. In this fact, the path loss propagation models have an important effect in Radio Mobile Systems and have found a favor in both research and industrial communities owing to their speed of execution.

The major focus of this review is based on earlier and present day developments encompassing the field of radio transmission and propagation. The leading aspect of this review involves an overall discussion of different models and techniques developed so far, facilitating radio propagation.

Keywords: path loss, models, attenuation

1. INTRODUCTION

Propagation model predicts the mean received signal strength for an arbitrary transmitter-receiver separation distances as well as the variability of the signal strength in a close spatial proximity to a particular location are useful in estimating the radio coverage area of a transmitter, since they characterize signal strength over large T-R separation distance (several hundreds or thousands of meter). On the other hand a Propagation model that characterizes the rapid fluctuation of received signal strength over very short travel distances are called small scale models.

As mobile moves over very small distances, the instantaneous received signal strength may fluctuate rapidly giving rise to small scale fading. The reason for this is that the received signal is a sum of many contributions coming from different directions. Propagation models are useful for predicting signal attenuation or path loss. This path loss information may be used as a controlling factor for system performance or coverage so as to achieve perfect reception. Propagation models are used extensively in network planning, particularly for conducting feasibility studies and during initial deployment. They are also very useful for performing interference studies as the deployment proceeds.

2. PROPAGATION MODELS

Basically Propagation Models are of two types

- A. Free Space Propagation Model
- B. Plane Earth Propagation Model

A. Free Space Propagation Model

The free space propagation Model is used to predict received signal strength when the transmitter and receiver have a clear and unobstructed line-of-sight path between them. In this kind of modeling the received power is a function of transmitted power, antenna gains & distance between transmitter and receiver. The logic behind it is that “the received power decreases as the square of the distance between transmitter & receiver. Major assumption of this kind of modeling is that there can only be single path without any obstruction between transmitter & receiver.

The following equation is employed for calculation of received power for distance of separation ‘I’ between transmitter & receiver.

$$P_r(I) = P_t G_t G_r \alpha^2 / 4\pi^2 I^2 L$$

Where,

- P_t is the transmitted power
- G_t & G_r are the transmitting & receiving antenna gains respectively
- L ($L \geq 1$) is the system loss & α is the wavelength of the concerned entity

The pathloss which represents signal attenuation as a positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power, and may or may not include the effect of the antenna gains. The path loss for the free space model when antenna gains are included is given by

$$PL(dB) = 10\log P_t/P_r = -10\log [G_t G_r \lambda^2 / 4\pi^2 I^2]$$

In a mobile radio channel, a single direct path between base station and a mobile is seldom the only physical means for propagation, and hence the free space propagation model is in most cases inaccurate when used alone.

B. Ground Reflection (Two-Ray) Model

The two ray ground reflection model is a useful propagation model that is based on geometric optics, and considers both the direct path and a ground reflected propagation path between transmitter and receiver. Here it is assumed that the total received power at the receiver end is the sum of powers due to two paths : first, the direct path between transmitter & receiver ;second, the path obtained by one ground reflection between the same transmitter & receiver separated by same distance as in case first. Another important parameter in this model is the height of location of receiver & transmitter with respect to the ground surface. The formula used for calculating received power at a distance ' I ' using this model is as

$$P_r(I) = P_t G_t G_r h_t^2 h_r^2 / I^4 L$$

where, h_t & h_r are the heights of locations of receiving & transmitting antennas with respect to the ground.

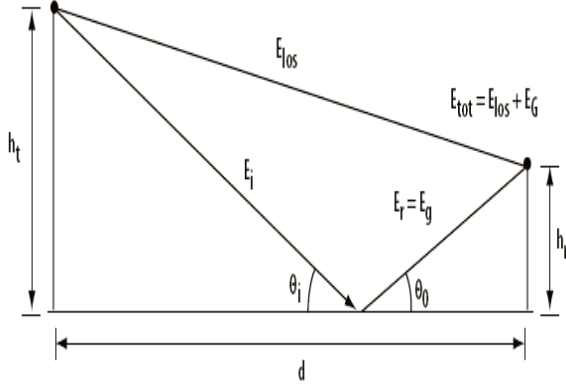


Fig.1 Depiction of Ground Reflection Two Ray Model

3. TYPES OF RADIO PROPAGATION MECHANISMS

The three very basic radio propagation mechanisms are actually the key to analysis of any radio propagation modeling based study. These are: reflection, diffraction & scattering.

Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings and walls.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities. The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when the line-of-sight path does not exist between transmitter and receiver.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel.

4. OUTDOOR PROPAGATION MODELS

Radio transmission in a mobile communication often takes place over irregular terrain. The terrain profile of an area needs to be considered for estimating the path loss. These models aim to predict signal strength at a particular receiving point in a specific local area. Most of these models are based on a systematic interpretation of measurement data obtained in the service area.

Some of the commonly used outdoor propagation models are:

A. Okumura Model

An empirical model developed by Japanese radio scientist Okumura as a part of extensive measurement campaign conducted in 1968. It is applicable for frequency range from 150 MHz to 1920 MHz but can be extrapolated up to 3GHz, equivalently deployed for a distance range of 1Km to 100Km.

Okumura discovered that a good model for path loss profile is a simple power law relationship, where exponent μ is a function of frequency & antenna height.

Path loss in Okumura model is expressed as

$$L_{50}(I) [\text{dB}] = L_F(I) + A_M(f, I) - G(h_t) - G(h_r) - G_{\text{area}}$$

where,

L_{50} : 50 Th percentile of path loss or median value.

$L_F(I)$: Free space propagation path loss.

$A_M(f, I)$: Median attenuation relative to free space.

$G(h_t)$: Base station antenna height gain factor.

$G(h_r)$: Mobile station antenna height gain factor.

G_{area} : Gain due to type of environment.

$G(h_r) = 20 \log(h_t/200)$ $1000\text{m} > h_r > 30\text{m}$

$G(h_r) = 10 \log(h_r/3)$ $h_r \leq 3\text{m}$

$G(h_r) = 20 \log(h_r/3)$ $3 < h_r < 10\text{m}$

H_t = transmitter antenna height.

H_r = receiver antenna height.

Okumura analyzed path-loss characteristics based on a large amount of experimental data collected around Tokyo, Japan. He selected propagation path conditions and obtained the average path-loss curves under flat urban areas. Then he applied several correction factors for other propagation conditions.

Okumura's Model is widely used for signal prediction in urban areas. It is considered to be simplest and best in terms of accuracy in path loss prediction in cluttered environment. The major disadvantage is its slow response to rapid changes in terrain; therefore model is good in urban and suburban areas but not as good in rural areas.

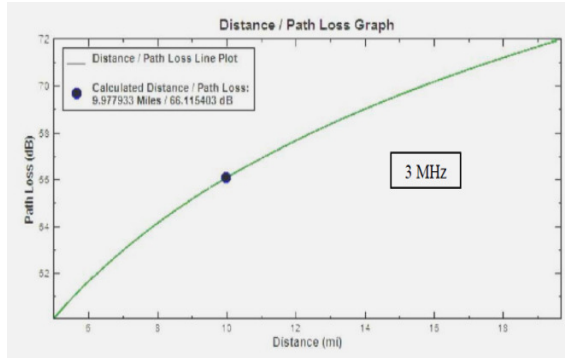


Fig. 2 Distance vs Path Loss Graph (3 MHz)

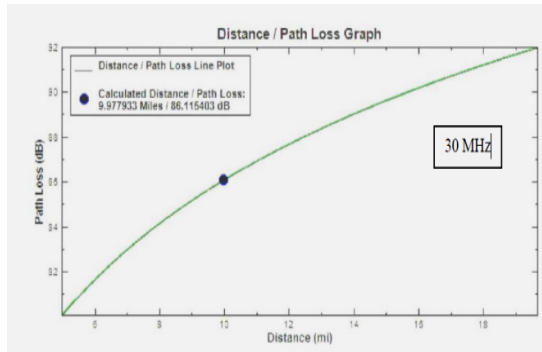


Fig. 3 Distance vs Path Loss Graph (30 MHz)

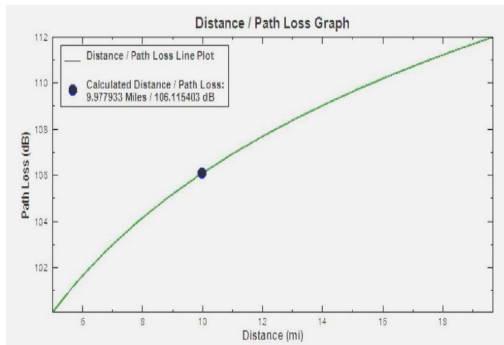


Fig. 4 Distance vs Path Loss Graph (300 MHz)

B. Hata Model

The Hata Model is an empirical formulation of the graphical path loss data provided by Okumura & is valid from 150MHz to 1500MHz. this model is well suited for large cell mobile systems but not for PCS which have cells of the order of 1km radius.

Median path loss for the Hata model is given by $L(\text{dB}) = 69.55 + 26.16\log(f_c) - 13.82\log(h_t) - a[h_r(m)] + (44.99 - 6.55\log(h_t)[m])[\log(I)]$

Where,

- f_c : frequency in MHz
- h_t : base station antenna height ranging from 30m to 200m
- h_r : mobile station antenna height.
- I : tx-rx separation [in Km].
- $A(h_r)$: antenna height correction factor for the mobile antenna as a function of coverage area

Hata model formula for path loss in urban area is given by

$$L_{50}(\text{dB}) = 69.55 + 26.16\log(f_c) - 13.82\log(h_t) - a(h_r(m)) + (44.99 - 6.55\log(h_t)[m])\log(I)$$

C. PCS EXTENSION TO HATA MODEL

The European Cooperative for Scientific & Technical research (EURO-COST) formed the Cost 231 model working committee to develop an extended and enhanced version of Hata Model and is valid from 1500-2000MHz.

The Cost231 median path loss is given by:

$$L_{50}(\text{dB}) = 46.3 + 33.9\log(f_c) - 13.82\log(h_t) - a(h_r) + [44.9 - 6.55\log(h_t)]\log(I) + C$$

where,

- F_c : is the frequency in MHz
- H_t : is the base station height in meters
- H_r : is the mobile station height in meters
- $a(h_r)$: is the mobile antenna height correction factor defined earlier
- I : is the link distance in km
- C : 0dB for medium cities or suburban centers with medium tree density
- C : 3dB for metropolitan centers

The Cost 231 model also called PCS Extension to HATA Model is used only for applications where the base station antenna is above certain roof tops.

D. Walfisch and Bertoni Model

This model is developed by Walfisch & Bertoni and it considers the impact of rooftops and building height by using diffraction to predict average signal strength at street level.

$$\text{Path loss, } S = P_0 Q^2 P_1$$

where,

- P_0 : free space path loss between isotropic antennas
- Q^2 : reduction in the rooftop signal due to row of buildings
- P_1 : signal loss from rooftop to the street

In dB, the path loss is given by

$$S \text{ (dB)} = L_0 + L_{\text{rts}} + L_{\text{ms}}$$

where,

- L_{rts} : rooftop to street diffraction and scatter loss
- L_{ms} : multiscreen diffraction loss due to rows of buildings
- L_0 : Free Space Loss

E. Longley-Rice Model

It is applicable to point-to-point communication systems in the frequency range from 40 MHz to 100GHz. The media transmission loss is predicted using the path geometry of the terrain profile and the refractivity of the troposphere.

The Longley-Rice Model operates in two modes

- a) When detailed terrain path profile is available, the path specific parameters can be easily determined and the prediction is called point to point mode prediction.
- b) If the terrain path profile is not available, the Longley – Rice method provides techniques to estimate the path specific parameters, and such a prediction is called an area mode prediction.
- c) One shortcoming of the Longley-Rice model is that it does not provide a way of determining corrections due to environmental factors in the immediate vicinity of the mobile receiver

5. CONCLUSION

From the survey of important models discussed, it gives a clear insight into the fact that for an analysis to be drawn out regarding specific model with respect to any data set the median values of attenuation or path loss is calculated

specific to the kind of environment, terrain & other such factors.

REFERENCES

- [1] Yih-Guang Jan, Yang-Han Lee, Hsien-Wei Tseng, Ming Hsueh Chuang, Jheng Yao Lin, Chih-Wei su, 'comparison of propagation path loss models', IEEE- C802.16m-07/121, call for comments, 2007.
- [2] V.S Abhaywardhana, I.J Wassell, D. Crosby, M.P Sellars, M.G Brown, 'comparison of empirical propagation path loss models for fixed wireless access systems' ,IEEE Conference, 2005.
- [3] Iskander M. and Yun Z., "Propagation Prediction Models for Wireless Communication Systems", IEEE Trans on microwave theory and Techniques, Vol. 50, No. 3, march 2002
- [4] Erceg, V. and et al, (1999) "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments" IEEE journal on selected areas in communications, vol. 17, no. 7.
- [5] Moinuddin A .A. and Singh S, "Accurate Path Loss Prediction in Wireless Environment", Institution of Engineers (India), Vol 88, July 2007, pp. 09 – 13
- [6] K. Allesbrook and J. D. Parsons, "Mobile Radio Propagation in British Cities at Frequencies in the VHF and UHF Bands, " IEEE Trans. on Vehicular Tech., Nov. 1977, pp. 313-323.
- [7] M. Barbiroli et al., "A new statistical approach for urban environment propagation modeling," IEEE Trans. on Vehicular Tech., Sept. 2002, pp 1234-1241
- [8] K. Bullington, "Radio Propagation for Vehicular Communications," IEEE Trans. on Vehicular Tech., Nov. 1977, pp. 295-308.
- [9] J. J. Egli, "Radio Propagation above 40 Mc Over Irregular Terrain," Proc. IRE, Oct. 1957, pp. 1383-1391.
- [10] M. Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services," IEEE Trans. on Vehicular Tech., August 1980, pp. 317-325. [Hata model]
- [11] M. F. Iskander and Z. Yun, "Propagation Prediction Models for Wireless Communication Systems," (Invited paper) IEEE Trans. on Microwave Theory and Techniques, Vol. 50, pp. 662-673 (March 2002).
- [12] T. K. Sarkar, et al., "A survey of various propagation models for mobile communication," IEEE Antennas and Propagation Magazine, Volume 45, Issue 3, June 2003, pp. 51-82.
- [13] M. W. Weiner, "Use of the Longley-Rice and Johnson-Gierhart Tropospheric Radio Propagation Programs: 0.02-20 GHz," IEEE J. on Selected Areas in Common, March 1986, pp. 297-307.
- [14] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation Measurements and Models for Wireless Communications Channels," IEEE Communications Magazine, vol. 33, pp. 42-49, January 1995.
- [15] A. Hills, et al., "Estimating Signal Strengths in the Design of an Indoor Wireless Network," IEEE Trans. Wireless Comm., vol. 3, pp. 17-19 (January 2004).

An Overview of Neural Filters for Impulse Noise Removal

Rashmi Kumari¹, S.K. Aggarwal²

^{1,2}ECE Deptt., Galgotias University ECE Deptt., JIT University
Greater Noida, U.P. Jhunjhunu, Rajasthan
¹rashmi167k@gmail.com ²aggarwal.ece.sk@gmail.com

Abstract: Impulse noise occurs frequently in image processing. This problem becomes important when the important details of images are corrupted and it is required to retrieve the lost details. Artificial neural network is used to play the vital role in function approximation. The restoration filter model for impulse corrupted digital images produces the better result using this noble technique. In this paper we are giving the overview of ANN in image processing and some existing neural network based filters for restoration of impulse corrupted images.

Index Terms: Artificial Neural Network, Backpropagation algorithm, Hopfield Model, Image processing, Impulse Noise

1. INTRODUCTION

Impulse noise removal is the basic preprocessing step in image processing as images are often corrupted during acquisition or transmission process [1]. Median filter is the most fundamental approach for the removal of impulse noise and several filters have designed by modifying the concept of median filter [2-5] due to its simplicity and capability to preserve some of the image details but the method to differentiate between the edge and corrupted pixel is still not appropriate. In the past two decades the concept of artificial neural network (ANN) is used in image processing and the outcome is comparatively effective in edge preservation and pattern related task.

In this paper we will discuss the ideas behind and the constructions of some neural based filters: Pulse coupled neural network (PCNN), Hopfield neural network (HNN), Resilient backpropagation algorithm and Radial basis function (RBFN). Our goal is to give an overview of these filters and the differences between them for restoration of impulse corrupted digital images. This paper is organized in this way that section II gives the idea about application of ANN in image processing and section III describes the method of popular ANN algorithms for restoration of impulse noise corrupted digital images.

2. ANN IN IMAGE PROCESSING

Artificial Neural Networks (ANN) can be viewed as weighted directed graphs in which artificial neurons are

nodes and directed edges (with weights) are connections between neurons. Use of neural network is increased as an alternative of pattern classifier and clustering. The Parzen window and the Bayesian discriminant were very popular as pattern recognition before ANN evolved as a powerful tool. The current role of ANN in image processing exceeds from the traditional applications. Feed-forward neural network is used for feature based segmentation and object recognition. Feed forward ANN, backpropagation and self organizing map (SOM) extends to lower level preprocessing task such as deblurring, denoising etc. Hopfield ANN is used in optimization related task and it gives the better solution for complex algorithms. For reconstruction of computerized tomography image modified Hopfield network is trained to perform the inverse Radon transform [6]. There are six steps of image processing task in which neural network is associated. Steps are as follows:

- Preprocessing : Noise Suppression, Deblurring, Image enhancement, Edge detection
- Data Reduction : Compression, Feature extraction
- Segmentation: Texture segregation, Colour recognition, Clustering.
- Object Recognition: Template matching, Feature based recognition
- Image understanding: Scene analysis, Object arrangement
- Optimization: Group matching, Automatic thresholding.

Here optimization is not a separate step but it supports to other steps of digital image processing. The performance of neural network algorithm depends upon the abstracted input data as pixel, local feature, edge, properties of individual objects etc. Image preprocessing operation generally falls into one of three categories:

- Image reconstruction
- Image restoration
- Image enhancement

The available neural network algorithms which are used for different preprocessing operations are categorized in Table 1. Adaptive resonance theory (ART) is employed for image enhancement whereas Hopfield can be applied for all three preprocessing category. The combination of two paradigms fuzzy logic and neural network (NF) is generalizing for all kind of image processing task now a days where there is a big probability of getting better results.

Table 1: ANN algorithms for preprocessing operations

Pre-processing category	Neural methods
Restoration	Feed Forward, Hopfield, PCNN, GANF, NF
Reconstruction	Hopfield, Adaline, NF
Enhancement	Hopfield, ART, NF

3. IMPULSE NOISE REMOVAL USING ANN

The majority of application of ANN in image processing is found in restoration process. A regression feedforward network in a convolution like way to suppress noise by using 5x5 pixel window as input is proposed by Greenhil and Davies [7]. Cellular neural network is proposed by Chua and Yang [8,9] for image processing in which all nodes are connected and each node contains a control template and a feedback template. Hanek and Ansari proposed a ANN architecture GANF (generalized adaptive neural filter) [10] for noise removal which is based upon stack filters that uses binary decomposition of grey value data. A network of networks architecture is proposed by Guan et al. for noise suppression [11]. Based on the connection architecture of ANNs, it can be grouped into two categories:-

Feed forward networks as: Single Layer Perceptron, Multilayer Perceptron, and Radial Basis Function.

Recurrent Networks as: Competitive Networks, Kohonen's Network, Hopfield Networks, and ART Models.

Here we are discussing the single layer model (PCNN), multilayer model (RBFN) then Hopfield model (HNN) as an example of recurrent network and resilient backpropagation method as a feedforward architecture for impulse noise detection and removal.

A. Single layer Model: PCNN

Pulse coupled neural network is based on the Eckhorn's model [12-13]. This model is developed on the basis of the visual properties of the mammal. The algorithm of PCNN is derived on the observation of the visual cortex nerve cell of cats and simulating the activities of the visual nerve cell. It is a single layer model suited for image segmentation, noise

reduction and image smoothness [14-15]. PCNN architecture is a composition of multiple nodes which forms a 2-D net. Single linking PCNN neuron's model is as shown in figure 1.

Mathematical equations are as below:

Feeding input of i th and j th neuron $F_{ij}(n) = S_{ij}$ (input impulse signal) (1)

Linking $L_{ij} = 1$ as it is a single linking network (2)

Internal activity $U_{ij}(n) = F_{ij}(n)(1 + \beta L_{ij}(n))$ (3)

Dynamic threshold $\theta = \alpha(1 + \beta L)$. (4)

Where α is a constant and its value lies within [0.96, 1] and β is linking strength. If α is set to 1 then the pixels having highest intensity will fire according to the value of threshold, so noise and signal can be separated.

$$\text{Output} = Y_{ij}(n) = \text{Step}[U_{ij}(n) - \theta] = \begin{cases} 1, & U_{ij}(n) \geq \theta \\ 0, & \text{Otherwise} \end{cases} \quad (5)$$

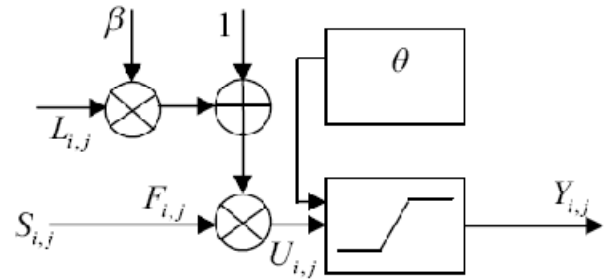


Fig. 1. Single linking PCNN neuron's model [16]

For impulse noise filtration, feeding input corresponds to the intensity scale of the image. Noise intensity is defined as the ratio of the number of highest or lowest intensity pixel and size of the noisy image. When the noise intensity is high the filtering procedure should be repetitive.

The neuron will fire first which intensity value is greater than the threshold and the output of all associated neurons will be affected simultaneously within the same region. The intensity value of the pixel and the output of the neighbor neurons is responsible for firing of other neurons and then output sequence $Y(n)$ is produced. After detection of impulse corrupted pixel the noisy pixel value is replaced by the value computed by adaptive median filter in the filtering window while other pixels remain unchanged.

B. Multi Layer Model: Radial Basis Function

Radial basis function: There are several structures available in ANN in which Radial Basis Function (RBF) is famous

one due to some important advantages over multilayer perceptron. The activation function of RBFN is generally a Gaussian function. There is no weight connected between the input layer and hidden layer. Standard Gaussian basis function is defined as:

$$\phi_i = e^{-\|x - c_i\|^2 / 2\sigma_i^2} \quad (6)$$

Where C_i and σ_j are parameters known as centre and global deviation that determines the spread of each basis function respectively. The set of pairs of input and desired output is decided by setting the values of parameters so that the training of network can be realized well. Training can be sequential or block type. The centre can be initialized by using K-means clustering algorithm. The architecture of RBFN used is shown in figure 2.

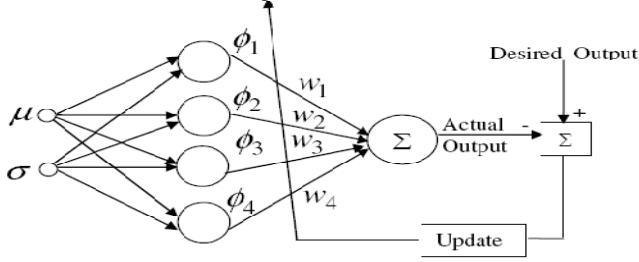


Fig. 2. Radial basis function network [17]

This detector has two layered structure. The input to this network is global standard deviation σ and mean μ of the noisy image. The difference of actual output and desired output gives error factor which is used to update the weight matrix using LMS algorithm depending on the value of error factor.

C. Recurrent Networks : Hopfield Model

The Hopfield model consists of a single layer of processing elements where each unit is connected to every other unit in the network other than itself. This structure is also called as self feedback or recurrent ANN. It takes binary value as inputs and gives binary value as outputs. The network is trained by input vectors or pattern vectors corresponding to different classes. Modified HNN [18] was used for edge detection in gray scale images which is using here for impulse noise detection.

Transmission rule is:

$$x_{t+1} = \text{sgn}[wx_t], \quad t = 1, 2, \dots, T \quad (7) \quad (8)$$

$$\text{sgn}[a_j] = \begin{cases} a_j = 1 & \text{if } a_j > 0 \\ a_j = -1 & \text{if } a_j < 0 \\ \text{Else} & a_j = 0 \end{cases}$$

Here the symbols have usual meaning as W stands for weight matrix, x is an input vector, a is an activation function and η is learning rate parameter. The learning rule is as:

$$W_{k+1} = W_k + \eta[x_0 x_0^T - (Wx_t)(Wx_t)^T] \quad (9)$$

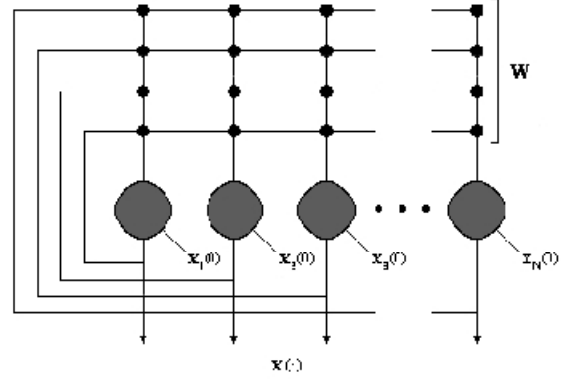


Fig. 3. Architecture of HNN unit [18]

The every pixel of the grey scale image is encoded by 8 bits (0-255) and it is rescaled first in the range of $[-1, +1]$. The 3×3 window mask gives input vector or 9-dimensional feature vector by which is set into binary values by using signum function after first iteration.

One centre pixel and 8 neighborhood gives 9 dimensional feature vector, so there are 9 nodes in HNN. If the centre pixel being considered is in a smooth region then all the 9 nodes behave in same way and output will be either +1 or -1.

If the centre pixel is corrupted then the output of the 8 neighborhood pixels will be in contrast of the centre pixel i.e., if value of output of neighborhood pixel is +1 the output of centre pixel will be -1 or vice-versa.

+1	+1	+1	+1	-1	+1	+1	+1	+1
----	----	----	----	----	----	----	----	----

OR

-1	-1	-1	-1	+1	-1	-1	-1	-1
----	----	----	----	----	----	----	----	----

Here a single output '1' denotes the noisy pixel and '0' denotes the non-noisy pixel. So a trained HNN by any test image separates the corrupted pixel and uncorrupted pixels.

$$Output = \begin{cases} 1 & \text{if } \sum_{i=1, i \neq 5}^9 x_i = 8 \text{ and } x_5 = -1 \text{ or} \\ & \sum_{i=1, i \neq 5}^9 x_i = -8 \text{ and } x_5 = 1 \\ 0 & \text{Otherwise} \end{cases} \quad (10)$$

D. Feedforward Network : Resilient Backpropagation

Sigmoid transfer function is used generally in the hidden layers of multilayer network, this is also called squashing function because it compresses as infinite input range into a finite output range. For the large value of input the slope must approach to zero in sigmoid function. This causes a problem when using steepest descent to train a multilayer network, since the gradient can have a very small magnitude and therefore, cause small changes in the weights and biases although the optimal value of these weights and biases have not been reached.

Resilient backpropagation training algorithm [20-21] is used to eliminate these harmful effects of the magnitude of the partial derivatives only the sign of the derivative is used to determine the direction of the weight update, the magnitude of the derivative has no effect on the weight update. The size of the weight change is determined by a separate update value factor, if the derivative of the performance function with respect to that weight has the same sign for two successive iteration, the value for each weight and bias is increased by a update value factor [22]. If the sign changes for two successive iterations then the value is decreased by a predetermined value factor, If the derivative is zero then the update value is unchanged.

$$\Delta w_{ji}(k) = \begin{cases} -A_{ji}(k) & \text{if } B(k) > 0 \\ +A_{ji}(k) & \text{if } B(k) < 0 \\ 0 & \text{else} \end{cases} \quad (11)$$

Where $B(k)$ is $\frac{\partial E}{\partial w_{ji}}(k)$ or derivative of error function $E(k)$ with respect to the W_{ij} and k represents the epoch index

$$A_{ji} = \begin{cases} \eta A_{ji}(k-1), & B(k-1)B(k) > 0 \\ \mu A_{ji}(k-1), & B(k-1)B(k) < 0 \\ A_{ji}(k-1), & \text{else} \end{cases} \quad (12)$$

Here $B(k-1)$ is $\frac{\partial E}{\partial w_{ji}}(k-1)$, η and μ are the increase and decrease factors, respectively and $0 < \mu < 1 < \eta$.

The original image is set as a target and the noisy image is set as an input which can be converted into column vector, training is done for the input-output pairs. The feedforward network can be designed as per the mask is chosen for input vectors. Finally the output vector is converted back into the matrix. This method gives the fastest learning way by which the detection of the impulse noise becomes very effective.

4. CONCLUSIONS

In this paper, we discussed the four different algorithms of ANN used in impulse noise removal. Either the algorithms are applied at the detection end or for the reduction process. The pair of original image and noisy image is used for training for detection of impulses and noisy image is taken as an input and noise free image is set as target image for reduction process. PCNN algorithms can also be designed for multilayer model but complexity of learning process will be increased whereas the Resilient backpropagation method is very fast as compared to others. The algorithms discussed above are having their own advantages and the combination of best suited methods for detection and reduction process can be applied on hit and trail basis to get the optimum output and for ease of complexity.

REFERENCES

- [1] Digital image processing, R. Gonzalez and R.Woods, PHI II Edition 2008
- [2] S-J. Ko and Y. H. Lee, "Centre-weighted median filters and their applications to image enhancement" IEEE Trans. Circuits and Syst., vol. 38, pp. 984-993, Sept 1991.
- [3] T. Sun and Y. Neuvo, "Detail-preserving median based filters in image processing," *Pattern Recognit. Lett.*, vol. 15, pp. 341-347, Apr. 1994.
- [4] Zhou Wang and David Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images," *IEEE Trans. Circuits & Systems II: Analog & Digital Signal Processing*, vol. 46, no. 1, pp. 78-80, Jan. 1999.
- [5] Brownrigg, D.: 'The weighted median filter', *Comm. Assoc. Comput.*, 1984, 27, (8), pp. 807-818.
- [6] V. Srinivasan, Y.K. Han, S.H. Ong, "Image reconstruction by a Hopfield neural network", *Image Vision Comput.*, vol. 11, 1993, pp. 278-282.
- [7] D. Greenhil, E.R. Davies, "Relative effectiveness of neural networks for image noise suppression", *Proceedings of the Pattern Recognition in Practice IV*, Vlieland, 1994, pp. 367-378.
- [8] W. Chua, L. Yang, "Cellular networks: theory", *IEEE Trans. Circuits Systems* vol.35,no.10,1988, pp.1257-1272.

-
- [9] W. Chua, L. Yang, "Cellular networks: applications", IEEE Trans. Circuits Systems, vol.35 ,no.10, 1988, pp.1273–1290.
 - [10] N. Ansari, Z.Z. Zhang, "Generalised adaptive neural filters", IEE Electron. Lett. 29 ,1993,pp.342–343.
 - [11] L. Guan, J.A. Anderson, J.P. Sutton, "A network of networks processing model for image regularization", IEEE Trans. Neural Networks, vol. 8 , no. 1, 1997, pp.169–174.
 - [12] R. Eckhorn, H.J.Reitboeck, M.Arndt, and P.W.Dicke PW, "Feature linking via synchronization among distributed assemblies: simulation of results from cat cortex", Neural Computation. Vol. 2, No. 3, 1990, pp.293-307.
 - [13] J.L. Johnsom, D. Ritter, "Observation of periodic waves in a pulse coupled neural network", Optics Letters. Vol. 18, No. 15,1993, pp.1253-1255.
 - [14] J. Zhang, J. Dong, M. Shi, "An adaptive method for image filtering with pulse coupled neural networks", Proceeding of International Conference on Image Processing, vol. 2, Genova, 2005, pp. 133- 136.
 - [15] Y. Ma, D. Lin, B. Zhang, C. Xia, "A novel algorithm of image enhancement based on pulse coupled neural network time matrix and rough set", Proceeding of Fourth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 3, China, 2007, pp. 86-90
 - [16] G.Cai, H.Li, D. Xu, H. Zhou , "Impulse noise filtering by using an adaptive single linking pulse coupled neural network"Procc. of IEEE conference, China, July 2010, pp.107-110.
 - [17] S. mohapatra, R. Dash, P.K.Sa, B. Manjhi, "Improved enhancement scheme using a RBFN detector for impulse noise, Procc. of ICTET, 2008, India, pp. 86-89.
 - [18] S. Chartier and R. Lepage, "Learning and extracting edges from images by a modified hopfield neural network", 16th International Conference on Pattern Recognition, vol.3, 2002, pp.431-434.
 - [19] M.A.A. Moustafa, "Qualitative and quantitative analysis of image enhancement techniques", IEEE midwest symposium on Circuits and systems, Cairo, 2005, pp. 664.
 - [20] Riedmiller, M., Braun, H., "A direct adaptive method for faster backpropagation learning: The Rprop algorithm", IEEE Int. Conf on Neural Networks, San Francisco, CA, 1993, pp.586-59 1.
 - [21] Riedmiller, M., "Rprop- Description and implementation details", Technical Report, University of Karlsruhe, Germany, 1994.
 - [22] E. Bedsok, M.Alci, "Using fast backpropagation algorithms for impulsive noise reduction from highly distorted images", IEEE midwest symposium on Circuits and systems, Cairo, 2005, pp. 940-943.

MOSFET's The New Generation Transistors

Rahul Gautam¹, Ankush Kapoor², Himanshu Saxena

^{1,2,3}Department of Electronics and Communication Engg.

Jawaharlal Nehru Government Engineering College, Sundernagar, Mandi, H.P. India

¹rahulgautam745@gmail.com, ²ankush8818@yahoo.com, ³hs.archdaemon@gmail.com

Abstract: Here we present a review of the major types of transistors used in the electronic industries, BIPOLAR JUNCTION TRANSISTORS and METAL OXIDE SEMICONDUCTOR FIELD EFFECT TRANSISTORS. Both the technologies are widely used and have their own importance but we will here present that how the use of the MOSFET's is much beneficial over the conventional BJT's. There are certain cases, as we will see, where the BJT's can completely be replaced by these new generation transistors: MOSFET's. Let us first get a brief overview to the basic three types of types of transistors.

1. INTRODUCTION

The advancement in technologies has far greater emphasized over the electronic era as large primitive circuits have shrunk to the size of few microns. The growth of the semiconductor technology, over the past few decades and drawn the attention of all the industries to its benefits.

For long, the bipolar technology ruled over the transistor industries. It had several advantages and continues to be used presently too. But there is always a chance for betterment. There are certain applications where the disadvantages of bipolar technology surpass its advantages. That is where the field-effect technology comes in handy. We shall review some of the advantages of the MOSFET's over the conventional BJT's. But first we shall introduce these types of transistors.

A. Bipolar Junction Transistor (BJT's)

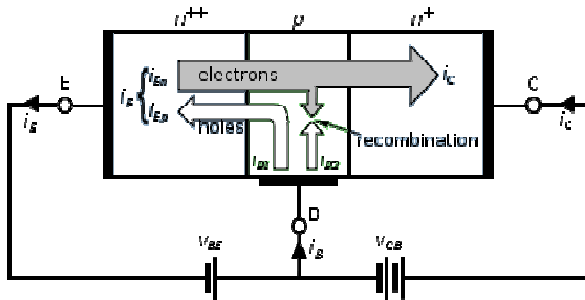


Fig. 1. Bipolar Junction Transistor (BJT)

Bipolar Junction Transistor or BJT was invented by Walter Brattain, John Bardeen and William Shockley in 1949 while working for Bell Telephone Laboratories.

A bipolar junction transistor consists of two types of semiconductors for its operation. A BJT can be used as amplifiers, switches or in oscillators. BJT's are so named because in them, the conduction takes place due to both, the majority and minority carriers. Charge flow in BJT's is due to bi-directional diffusion of charge carriers across a junction between two regions of different charge concentrations. BJT has three regions i.e. emitter, collector and base.

Emitter region is highly doped so as to regulate the flow of majority carriers through the emitter-base junction. Base region is thin and is of opposite polarity to the emitter and collector. Collector region is thick so as to accommodate charge carriers entering from base-collector junction.

BJT's have two types, i.e. PNP and NPN based on doping of the terminals of BJT.

B. Field Effect Transistor (FET's)

The field effect transistor is a transistor that uses an electric field to control the shape and hence the conductivity of a channel of one type of charge carrier in a semiconductor material. FET's are unipolar transistors as they involve single carrier for their operation.

FET's are majority charge carrier devices. The device consists of an active channel through which majority charge carriers (electrons or holes) flow from source to drain. Source and drain terminals are connected to the semiconductor through ohmic contacts.

The FET controls the flow of electrons or holes from the source to drain, by affecting the size and shape of a conductive channel, created and influenced by voltage applied across the gate and source terminals. This conductive channel serves as the stream through which electrons flow from source to drain.

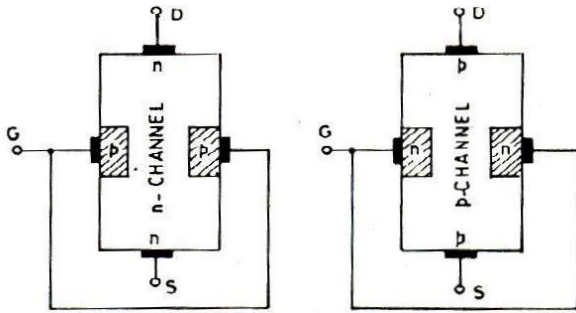


Fig. 2. - n-channel FET(left) p-channel FET (right)

C. Metal Oxide Semiconductor Field Effect Transistors (MOSFET's)

MOSFET is a unipolar transistor, which acts as a voltage-controlled current device i.e. current at two electrodes, viz. drain and source is controlled by the action of an electric field. This electric field is applied at another electrode gate having semiconductor and metal very a thin metal oxide layer in-between.

MOSFET's are the transistors used for amplifying or switching electronic signals. It is a four terminal device with source, gate drain and body. The body terminal is connected to the source terminal, thereby making it a three terminal device .

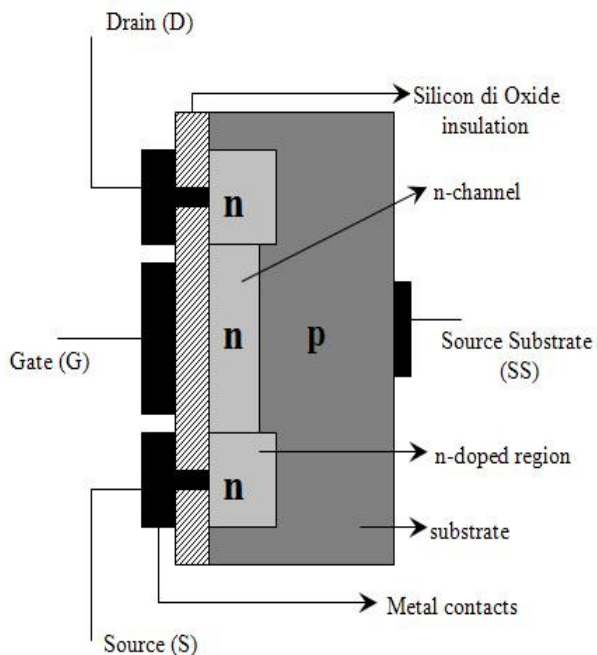


Fig. 3 n-channel Enhancement type MOSFET

MOSFET'S are classified on the basis of polarity of source and drain terminals i.e. if source and drain terminals are n-

type then the MOSFET is said to be n-MOSFET, and if they are of p-type, then the MOSFET is said to be p-MOSFET.

2. DRAWBACKS OF BJT'S

1. A BJT has a very complex base control. Hence, it may lead to confusion and as such, a BJT requires quite skillful handling.
2. The switching is not very fast compared to the high alternating frequency of voltage and current source. Often, we are required to operate with very high frequencies. In such cases, the normal Bipolar-Junction transistors are not effective in use as they do not give optimum response.
3. There is a problem of the thermal runaway in the BJT's. Since the collector has to dissipate more heat, therefore the transistor heats up and it may lead to the damage.
4. A BJT requires limiting voltage of 0.3V for the lowest voltage drop on the current path.
5. A BJT can only be used to amplify the current of the signal with current as it is a current controlled device.
6. BJT's, when working on high frequency produce large amount of noise while amplifying the signal.

3. ADVANTAGES OF FET'S OVER BJT'S

1. Field effect transistors have very high input impedance as compared to conventional bipolar junction transistors.
2. FET's are unipolar i.e. only majority charge carriers are responsible for conduction whereas in BJT's both majority and minority charge carriers are responsible for conduction.
3. FET's can be miniaturized to a very large extent as it requires only one type of charge carriers for conduction.
4. FET's are lesser noisy than BJT's, i.e. they produce less distortion in input signals while amplification, when employed in amplifiers.
5. FET's are voltage controlled devices i.e. they amplify voltage of a signal with voltage whereas BJT's are current controlled devices i.e. they amplify current with current.
6. BJT & FET parameters are temperature dependent. In BJT the collector junction resistance decreases (collector current increasing) with a temperature rise

due to the high temperature & the transistor will damage quickly. In FET drain resistance increasing (drain current decreasing) with increasing temperature. Due to this property it will not damage easily. We can say from the above two statements FET is more temperature stable. So, FET can use in high temperature applications.

4. ADVANTAGES OF USING MOSFET'S

- There is almost no saturation of the device, hence very small amount of time is elapsed for attaining the saturation, and hence the response time of the MOSFET's is lesser than the BJT's. Hence, the MOSFET's are much faster than the BJT's. This is quite useful in present era, when the speed of the operation is of utmost importance. The MOSFET's can furnish the operational frequencies up to 10 GHz with the transient speed 10-100ns because of almost no saturation.
- MOSFET's don't need current on their control pin, but require more voltage. Some don't turn on completely at 5V, some do. A BJT requires limiting voltage of 0.3V for the lowest voltage drop on the current path.
- MOSFET's are usually more efficient switches for power supplies, etc. where you want a switch rather than an amplifier. As has been seen that almost no saturation is involved in the MOSFET's, hence, they prove to be more suitable than the BJT's, for switching applications.
- MOSFET's have higher input impedance than BJT's. The input impedance is a measure of the resistance of input terminal of the transistor to electrical current. When designing voltage amplifiers it is desirable for the input resistance to be as high as possible. Therefore, MOSFET's are more widely used in the input stage of voltage amplifiers.
- MOSFET's can be made much smaller than BJT's. Many more MOSFET's can be placed in a smaller area than BJT's. For this reason, MOSFET's form bulk of the transistors used in microchips and computer processors. MOSFET's are also easier to manufacture than BJT's because MOSFET's involve fewer steps in their manufacturing process.
- MOSFET's are lesser noisy than BJT's. In an electronic context, noise refers to random interference in a signal. When a transistor is used to amplify a signal the internal processes of the transistor will introduce some of this random interference. BJT's generally introduce more noise into the signal than MOSFET's. This means MOSFET's are more suitable for signal processing applications or for voltage amplifiers.
- BJT's suffer from a problem known as "thermal runaway." Thermal runaway happens because the conductivity of a BJT increases with temperature. Because transistors tend to heat up in proportion to current flowing through them, this means that the conductivity and temperature of BJT's can increase exponentially. This can damage the BJT and makes designing circuits for BJT's more difficult. MOSFET's do not suffer from thermal runaway.
- A FET is a voltage controlled current device, which means a voltage turns it on (so we only have to worry about leakage current) and in essence it amplifies a voltage with a current. A BJT is a current controlled current device, so it amplifies a current with a current.
- When not in operation, a BJT still has some current flowing through it, which leads to some power loss. But this drawback is readily overcome in a MOSFET. As such, in the OFF state, the MOSFET is not having any current (in enhancement mode), which leads to the lower losses and protection of the circuit.
- Also, when the MOSFET's are employed in the amplifiers, it results in greater bandwidth, as compared to the ones employing the BJT's. It is because the MOSFET's do not involve any junction capacitances so the bandwidth increases.
- From integration point of view MOS transistors provide higher packing density compared to BJT. So IC's made by MOS transistor can be smaller.
- N Channel MOSFET'S need +2 to +4 V to turn them on. The gate current is approximately zero. But in BJT's base current starts to flow with an input voltage of about +0.7V. Relatively larger base currents are needed to make transistors operate. Hence the devices using the MOS technologies are much more sensitive, than the ones using the BJT's.
- In the digital integrated circuits also, the MOS technology has far more benefits than the bipolar technology. MOS logic has a greater fan-out of the gate. Also, it reduces the power consumption, which is very much required in large scale integration.

5. CIRCUIT SIMULATIONS:

Now we shall present some of the circuits using above three types of transistors. The simulation software that we shall be using is Toolkit for Interactive Network Analyses (TINA).

Following circuits show the basic 2-stage amplifiers employing BJT's, FET's and MOSFET's. Following are circuits and their frequency responses of the given circuits.

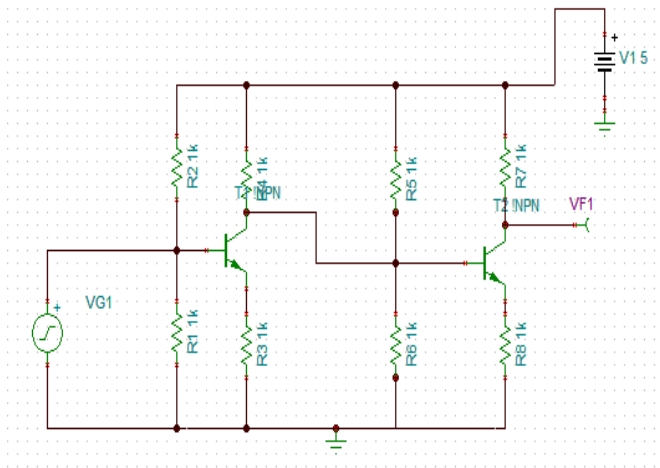


Fig.-A two stage Direct coupled amplifier using BJT's

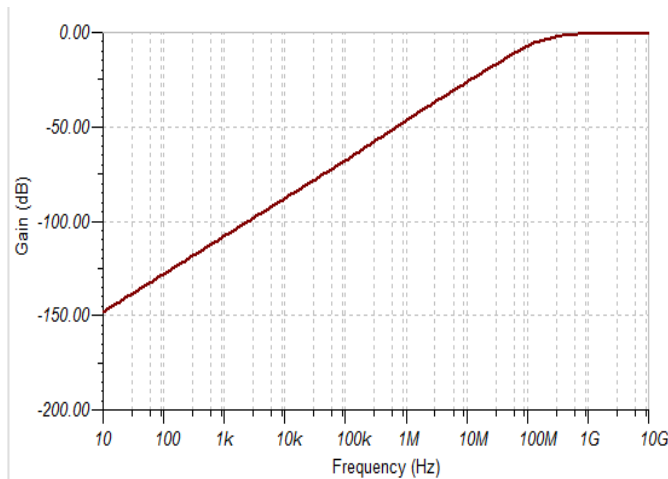


Fig.-Frequency response of a two stage direct coupled amplifier using BJT's

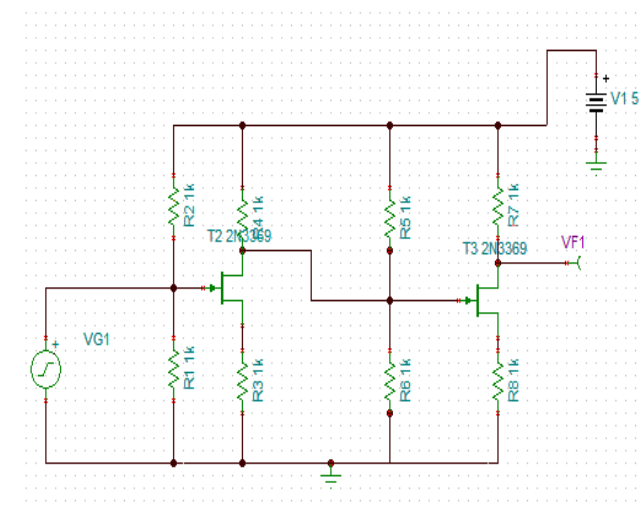


Fig.- A two stage direct coupled amplifier using FET's

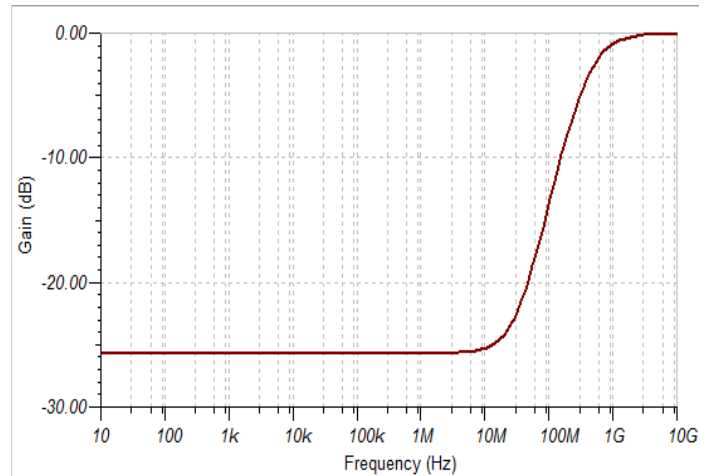


Fig.-Frequency response of a two stage direct coupled amplifier using FET's

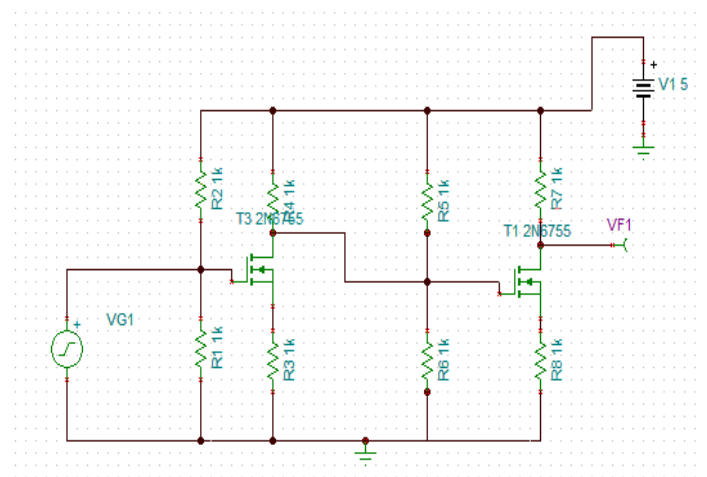


Fig.- A two stage direct coupled amplifier using MOSFET's

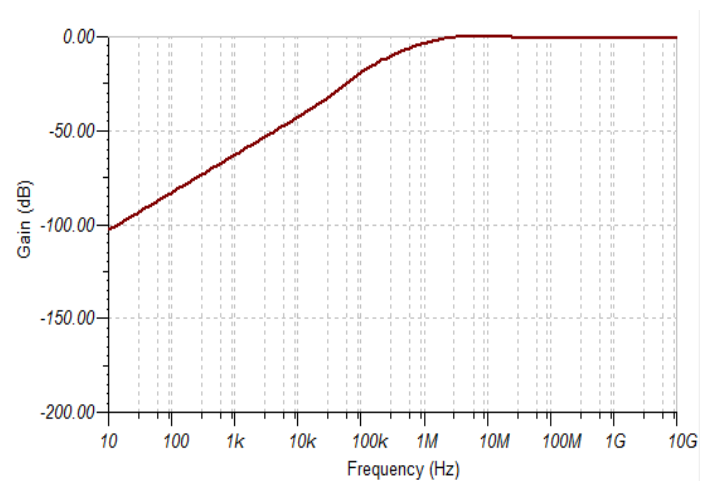


Fig.-Frequency response of a two stage direct coupled amplifier using MOSFET's.

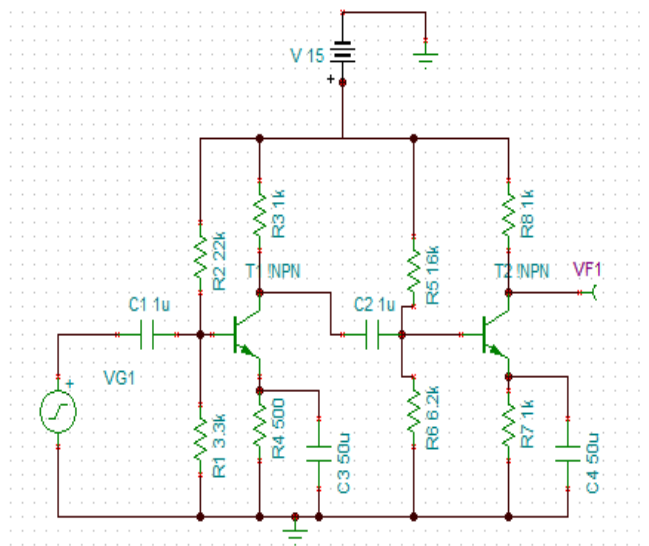


Fig. A two stage RC-Coupled amplifier using BJT's

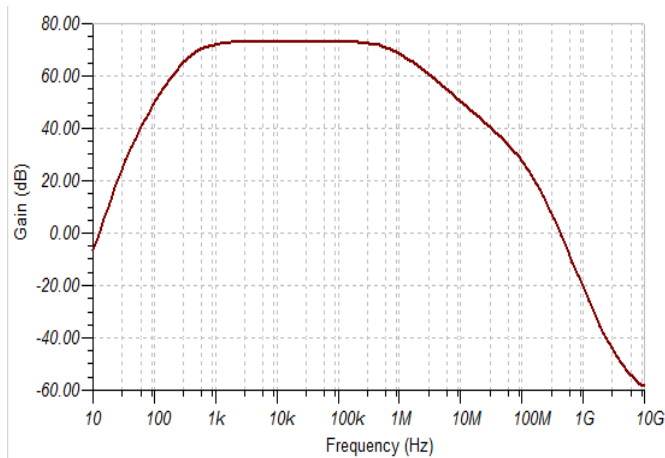


Fig.-Frequency response of a two stage RC-Coupled Amplifier using BJT's

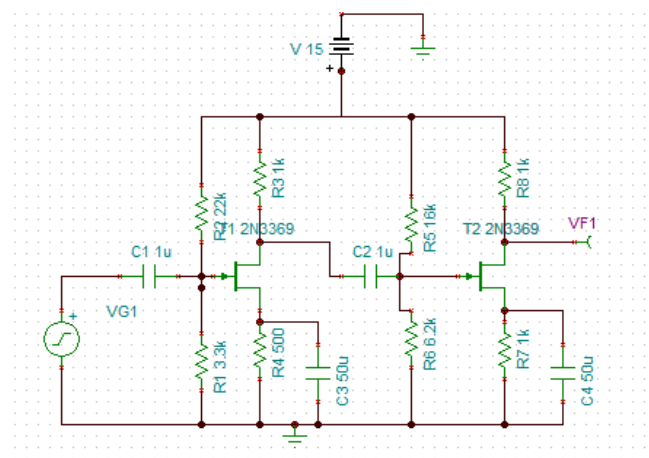


Fig.- RC-Coupled Amplifier using FET's

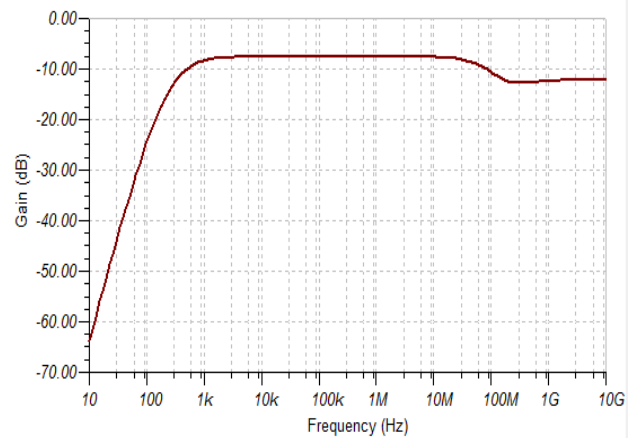


Fig.-Frequency response of a two stage RC coupled Amplifier using FET's.

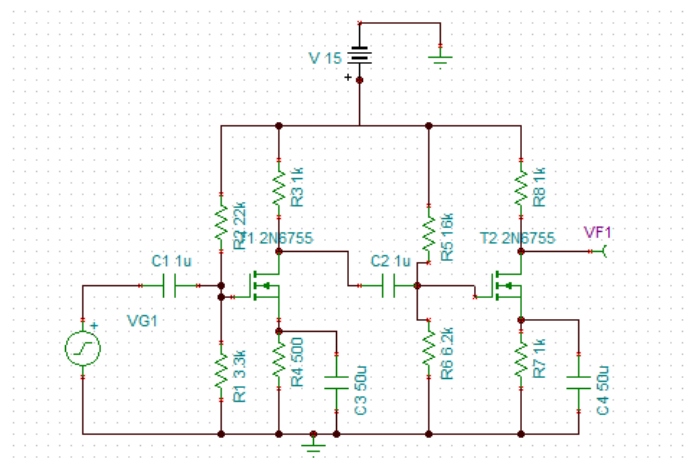


Fig.- A two stage RC Coupled Amplifier using MOSFET's

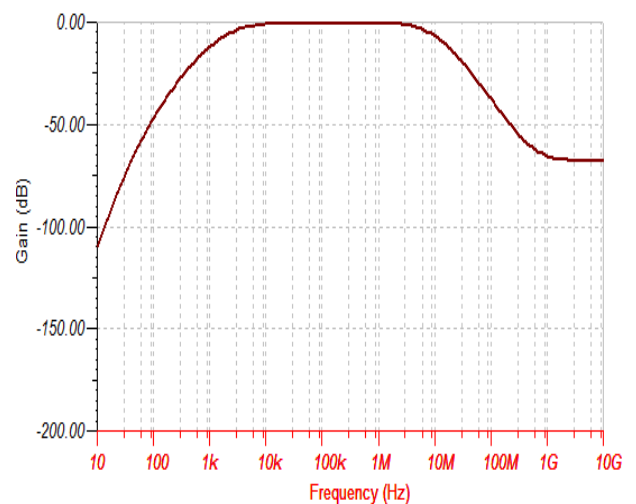


Fig.-Frequency response of a two stage RC Coupled Amplifier using MOSFET's.

6. COMPARISON

S. No.	CIRCUIT IN WHICH EMPLOYED	FREQUENCY RESPONSE OF		
		BJT's	FET's	MOSFET's
1.	Direct coupled 2-stage amplifier.	Gain increases linearly with frequency and then becomes constant.	Gain, first constant for low frequencies (up to 10MHz), then starts rising and ultimately becomes constant.	Gain is higher than BJT's, and is constant for a wider range of frequencies, so bandwidth of operation is larger.
2.	RC-Coupled 2-stage amplifier.	Gain first increases and then becomes constant and then falls due to inter-electrode capacitive effect.	The gain rises in the beginning, and then becomes constant for a fairly large range of frequencies, and then falls, but by a small amount.	Gain follows a similar trend but the variation is very less compared to the other two types of transistors, so bandwidth is increase practically.

7. CONCLUSION

With all that we have put together so far, we can say that the basic need of an electronic device is served fully by a BJT. But, always there is a possibility of betterment, in any technology. MOS technology is very much beneficial, for the reasons we have seen, and I presume that there is yet a lot to be developed in this field of transistor fabrication.

With the MOS technology in hand, we can achieve great heights in the run to achieve the reduction in size and increase in the speed of operation. With the continuous efforts of the scientists, engineers and researchers, this technology, I believe, is going to touch the sky in the field of electronics.

REFERENCES

- [1] J. D. Ryder, "Electronic fundamentals and applications".
- [2] N. N. Bhargava and Kulshrestha, "Electronic Devices".
- [3] J. Millman and C. C. Halkias, "Electronic circuits and devices".
- [4] Multiple-Gate SOI MOSFETs: Device Design Guideline, by Jong-Tae Park, Member, IEEE, and Jean-Pierre Colinge, Fellow, IEEE.
- [5] Use of nano-scale double-gate MOSFETs in low-power tunable current mode analog circuits, by Hesham F. A. Hamed Savas Kaya, Janusz A. Starzyk
- [6] Performance and reliability Design Issues for Deep-Sub micrometer MOSFET's, by James E. Chung, Min-Chie Jeng.
- [7] Scaled Silicon MOSFET's: Degradation of the Total Gate Capacitance, by Dragica Vasileska, Dieter K. Schroder.
- [8] Silicon on Insulator MOSFET Development from Single Gate to Multiple Gate, by Prashant Mani, Manoj Kumar Pandey.
- [9] Symmetric and Asymmetric Double Gate MOSFET Modeling by H. Abebe, E. Cumberbatch, H. Morris, V. Tyree, T. Numata, and S. Uno.
- [10] Power MOSFET Thermal Instability Operation Characterization Support, by John L. Shue and Henning W. Leidecker.
- [11] Theory of Ballistic Nanotransistors, by Anisur Rahman, Jing Guo, Supriyo Datta.
- [12] Essential Physics of Carrier Transport in Nanoscale MOSFETs, by Mark Lundstrom and Zhibin Ren.
- [13] Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits, Naveen Verma, Joyce Kwong, and Anantha P. Chandrakasan.
- [14] The Invariance of Characteristic Current Densities in Nanoscale MOSFETs and Its Impact on Algorithmic Design Methodologies and Design Porting of Si (Ge) (Bi) CMOS High-Speed Building Blocks, by Timothy O. Dickson, Kenneth H. K. Yau, Theodoros Chalvatzis, Alain M. Mangan, Ekaterina Laskin, Student Member, IEEE, Rudy Beerkens, Paul Westergaard Mihai Tazlauanu, Ming-Ta Yang, and Sorin P. Voinigescu

Integrated Circuits and MicroElectroMechanical Systems Fabrication A Review on Pre-existing Fabrication Techniques

Himanshu Saxena¹, Rahul Gautam², Ankush Kapoor³

^{1,2,3}Electronics and Communication Department,

Jawaharlal Nehru Government Engineering College, Sundernagar, India

¹hs.archdaemon@gmail.com, ²rahulgautam745@gmail.com, ³ankush8188@yahoo.com

Abstract: The continuous development in the field of Integrated Circuits so as to shrink electronic devices along with simultaneous increase in their performance has affected technology to a large extent. MicroElectroMechanical Systems (MEMS) are those integrated circuits which can only be appreciated with the help of microscope. As these MEMS is being used for various applications such as pressure sensors, accelerometers, and ink-jet print-heads. It is of prime importance to know how we can implement various fabricating techniques for the fabrication of MEMS.

IndexTerms: Integrated Circuits, Fabrication, MEMS, Microelectromechanical Systems

1. INTRODUCTION

Integrated circuits or the ICs are a low cost miniature electronic circuit which consists of various active and passive components which are connected to each other on a single silicon chip. These integrated components are different from conventional components in appearance but have similar working [1].

Integrated circuits can be classified into two broad categories:

- Monolithic Integrated Circuits, and
- Hybrid Integrated Circuits.

The word 'monolithic' is derived from Greek words 'monos' meaning single and 'lithos' meaning 'stone'. Therefore, those integrated circuits in which the components are fabricated on a single silicon chip. This type of integration is used in places where we require large number of similar chipsets.

Hybrid Integrated Circuits are those integrated circuits in which separate component parts are attached to a ceramic substrate and are interconnected by means of either wires or metallization. This type of fabrication is mostly used in places where we require small quantity of custom designed circuits.

MicroElectroMechanical Systems (MEMS)

MEMS contain mechanical elements that are built on such a small scale that they can be appreciated only with the help of microscope. MEMS elements interface with non-electronic signals and often merge signal processing with sensing. MEMS may contain mechanical parts, such as pressure sensors, flow sensors, or optical-beam handling devices. Some fully integrated MEMS are designed using computer-aided design (CAD) techniques based on VLSI and mechanical CAD systems; they are batch-fabricated using VLSI-based fabrication tools. Like VLSI, MEMS are becoming progressively smaller, faster, and more functional [2].

2. VARIOUS METHODS FOR IC FABRICATION

A. Solid State Diffusion

In solid state diffusion the impurity atoms replace silicon atoms in the lattice and are termed as substitutional impurities. Solid state diffusion consists of following steps [3]:

1). Predeposition step

In this step the impurities that are supposed to be diffused in the crystalline silicon chip are placed on the surface of the chip.

2) Diffusion step

The sample is then subjected to high temperature of about 1100°C for about an hour. Due to this high temperature the impurities diffuse into the crystalline silicon chip. Thereby, form p-n junction on the chip. The depth of this junction varies from 0.1µm to 20µm.

B. Photolithography

Photolithography is derived from *Latin* words 'photo' meaning 'light', 'lithos' meaning 'stone' and 'graphy'

meaning 'writing'. It is an optical technique for transferring patterns onto a crystal silicon chip.

Photolithography consists of various processes as described below [4]:

1) Oxidation of Crystalline Silicon

In this process the crystalline silicon is placed in an oxidizing environment and a layer of SiO_2 is deposited on the surface of crystalline silicon chip so as to prevent impurities to diffuse with the crystalline silicon chip. This helps in designing any component at a particular location on the surface of crystalline silicon chip by removal of oxide layer and then by diffusion of impurities at that particular location. It is usually $0.2\mu\text{m}$ to $1\mu\text{m}$ thick.

2) Coating with Photoresist

After the oxidation of crystalline silicon chip and the formation of SiO_2 layer it is coated with a thin layer of photosensitive material called photoresist. When this layer is exposed to a particular wavelength of light it undergoes chemical change and gets dissolved.

3) Masking

So as to select a particular area on the chip for any component photomask is used. It is an opaque sheet which prevents the surface of photoresist to be exposed to light. Light of appropriate wavelength is incident and the area with exposed photoresist gets dissolved.

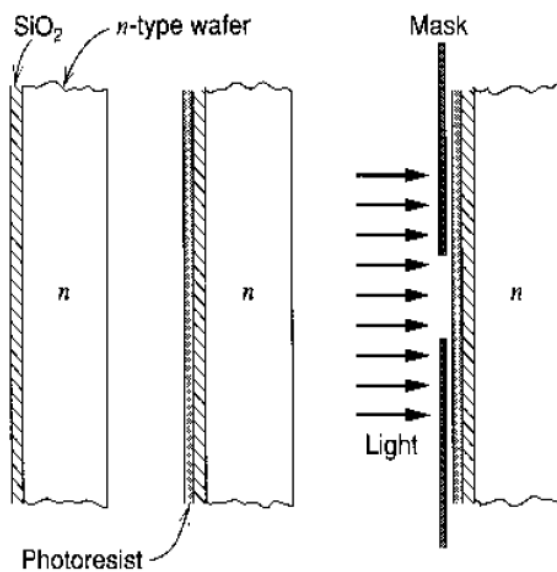


Fig. 1 Photolithography¹
Oxidation of crystalline silicon (left), Coating with photoresist (center), Masking (right)

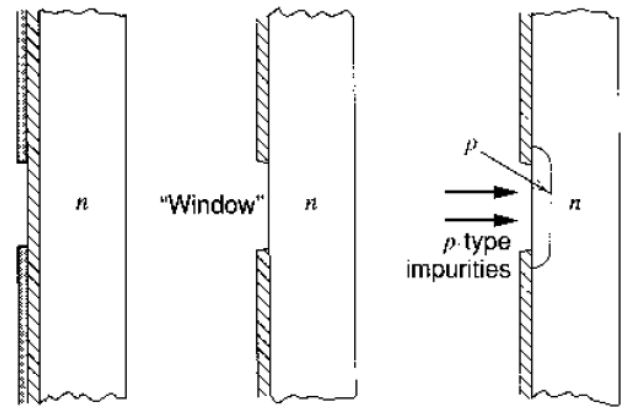


Fig. 2 Photolithography¹
Etching (left), Removal of remaining photoresist (center), Diffusion (right)

4) Etching

Now our aim is to remove the oxide layer from the surface of crystalline silicon chip. This is done by dipping it into a solution of hydrofluoric acid or exposing it to electrically produced plasma produced in a plasma etcher. After this process we are left with bare silicon surface.

5) Removal of remaining photoresist

The remaining photoresist is removed by the help of chemical stripping, leaving the samples with holes or windows at desired locations.

6) Diffusion

In the holes or the windows created, impurities are diffused so as to create a region with different majority carriers for example, if the crystalline silicon chip is of n-type semiconductor, p-type impurities are diffused so as to form a p-n junction.

C) Various Photolithography Techniques

1) Optical Lithography Techniques

Optical lithography comprises the formation of images with visible radiation in a photo resist using proximity or projection printing. These methods rely upon a mask to form the beam for the necessary image to be formed on the resist.

a) Proximity Optical Lithography

Proximity optical lithography is a relatively simple technique. It requires no image formation between the mask and the photoresist. The proximity optical lithography system consists of a light source, a mirror, a shutter, a mask and the stage on to which the photoresist is positioned [5].

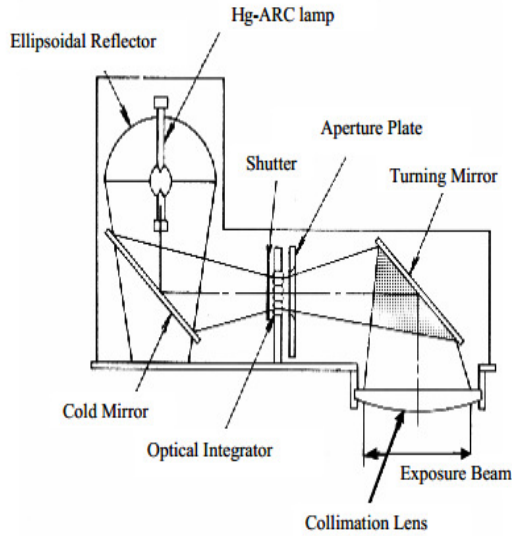


Fig. 3 System for proximity optical lithography⁵

In this technique the mask and the photoresist separation is usually around 20 μm to 50 μm . This leads to an acceptable resolution of around 500nm. As diffraction occurs between the mask and the photoresist, so resolution can be increased by narrowing the mask and resist separation.

b) X-Ray Lithography

The wavelength of incident radiation is decreased from visible to X-Ray spectrum. It offers a viable alternative to optical lithography as the requirement for smaller circuit patterns increases beyond the present capabilities of optical lithography. Using ultra-violet wavelengths of approximately 0.1nm gives a resolution which is significantly better than the optical system [6].

But with the very nature of X-Ray radiation the majority of absorbing materials at optical wavelengths such as chromium becomes transparent at X-Ray wavelengths.

In such lithographic technique an X-Ray source illuminates a mask which casts a shadow on the resist. Overlaying the pattern, made out of X-Ray absorbing material, on to a transmitting material, produces the mask used in X-Ray lithography. Any materials to be used in the absorbing part of the mask must have high absorption coefficients in the X-Ray region (The absorption coefficient of any elementary material of atomic number Z and density ρ is proportional to $\rho Z^4 \lambda^3$ over a wide range of wavelengths).

The mask casts shadows of the pattern on to the wafer below, the diffraction that may occur can be neglected

because the wavelength of the exposing radiation is so short compared to the optical case.

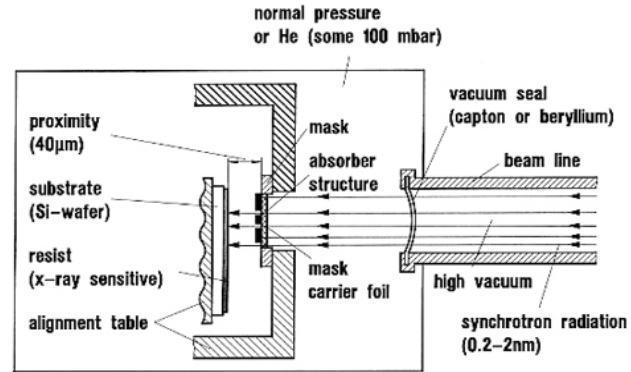


Fig. 4 System for proximity optical lithography⁷

However, the pattern cast by the radiation is not perfectly sharp due to the finite size of X-Ray source and cannot be considered as a point source at the mask. We must also take into consideration the fact that no material is entirely transmissive or absorptive, and therefore the intensity of the X-Rays will drop whilst passing through the mask.

While proximity lithography is the preferred technique when using X-Rays. However, the difficulties with using this technique are even more of a problem at X-Ray wavelengths than at ultra-violet wavelengths. The materials which form the multi-layer mirrors have to be strong absorbers at short wavelengths, and the layers themselves must be only a few nanometers in thickness, hence being more difficult to produce than their ultra-violet lithography counterparts.

We must consider projection X-Ray lithography as very much a technique under development, and one that may become a useful tool in the future. We know from optical lithography that the resolution will be better than the proximity technique, but the manufacturing technique for the mirrors needs to be enhanced before projection can become a viable alternative in the X-Ray region.

2) Particle based lithography

a) Electron beam lithography

When we require resolution much less than that given by X-Ray lithographic techniques, from the quantum mechanical principle of wave-particle duality and the de Broglie equation ($\lambda = h/(2me\Delta V)^{1/2}$), we find that an electron with energy of 10keV has a wavelength of around 12pm. This obviously represents a huge reduction in wavelength compared to X-Ray radiation, and therefore electron considered [7].

Electron beam lithography replaces the photons with an electron beam, and utilizes a different optical system with image formation between the source and the resist with no mask in the system because we are using a beam of electrons whose direction can be controlled by electrical and magnetic field.

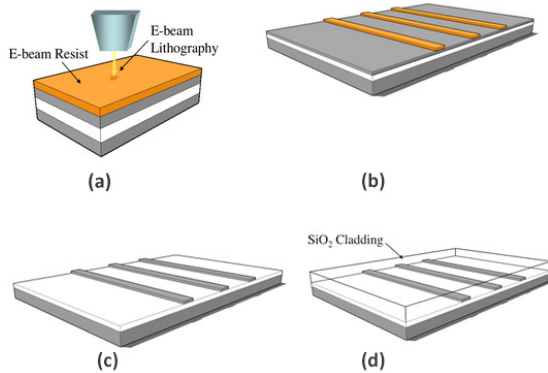


Fig. 5 a) Photonic structures are defined by E-beam lithography, b) The resist is developed leaving the resist mask, c) Reactive ion etching is used to etch the mask pattern into the top silicon layer, down the buried oxide layer, and d) This device is then covered with a cladding of oxide using plasma enhanced deposition.⁸

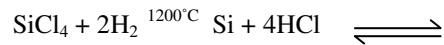
In this process a sample of oxidized silicon wafer is coated with negative electron beam resist that becomes insoluble in developing solutions when exposed to radiation. The electron beam is then incident on the surface of silicon wafer which is used to define photonic structures (Figure 5a). The resist is then developed leaving the resist mask (Figure 5b). Then the etching process is done using Reactive Ion Etching (Figure 5c). The device is covered with a cladding of silicon oxide using Plasma Enhanced Chemical Vapor Deposition (Figure 5d) [8].

Writing very fine patterns using this technique on the resist is a very slow process as only one point on the resist is exposed at any given time. There has been considerable interest in improving the throughput in electron beam lithography, generally concentrating on shaping and enlarging the size of the beam.

C) Epitaxial Growth

Epitaxial growth has been derived from Greek word ‘epi’ meaning ‘upon’ and ‘teionon’ meaning ‘arranged’. Thus, it can be described as the process of arranging atoms in a single crystal fashion upon a single crystal substrate such that the resulting layer is extension of the substrate crystal structure [9].

It involves the hydrogen reduction of silicon tetrachloride.



An epitaxial film with specific concentration of impurity is required. Hence, phosphine (PH_3) for the n-type and diborane (B_2H_6) for p-type doping into the silicon-tetrachloride hydrogen gas stream.

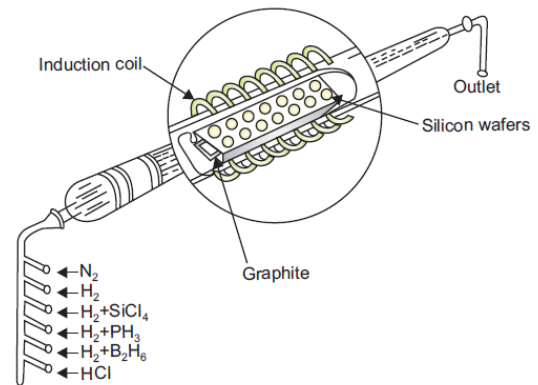


Fig. 5 System for growing epitaxial film²

The process of epitaxial growth is carried out in a reaction chamber consisting of long cylindrical quartz tube encircled by an RF induction coil. The crystalline silicon wafers are placed on a rectangular graphite rod called boat.

This boat is then placed in the reaction chamber where the graphite is heated inductively to a temperature 1200°C . The various gases required for the growth of desired epitaxial layers are introduced into the system through a control console.

D) Ion Implantation

The concept of ion implantation was given by William Shockley in 1956. But the first implanters were introduced in the markets after 1973.

Basically, it is a low temperature technique. In ion implantation, first, the dopant atoms are volatilized. Then they are ionized, and then accelerated to high velocities, by making use of high electric fields, so that now they are having very high energies. Now, these accelerated ions are separated by the mass-to-charge ratios.

Finally, these high energy ions are directed at the silicon wafer (substrate). These atoms now enter the crystal lattice and there occurs a collision between the high energy atoms and the host atoms due to which the impurity atoms lose energy, and finally come to rest after penetrating through some depth within the solid. To what depths do the dopant atoms penetrate the solid depends upon the dopant, substrate materials, and acceleration energy [10].

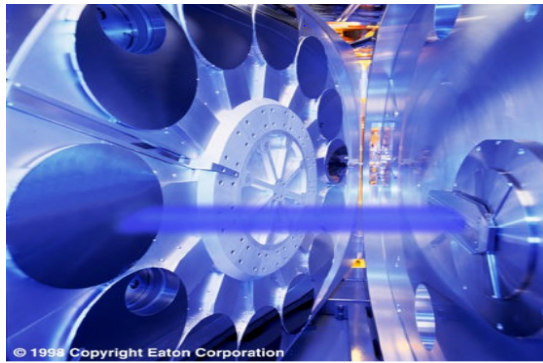


Fig. 6 Eaton HE3 High Energy Implanter, showing ion beam hitting the 300mm wafer end-station.³

E. Local Oxidation

The need for local oxidation arises when we want to fabricate the regions of the silicon surface that are covered with a thin layer of silicon dioxide. Also, sometimes we require the transition from thick to thin regions without introducing a large vertical step, so that the other processes, like metallization, are on a relatively planar surface.

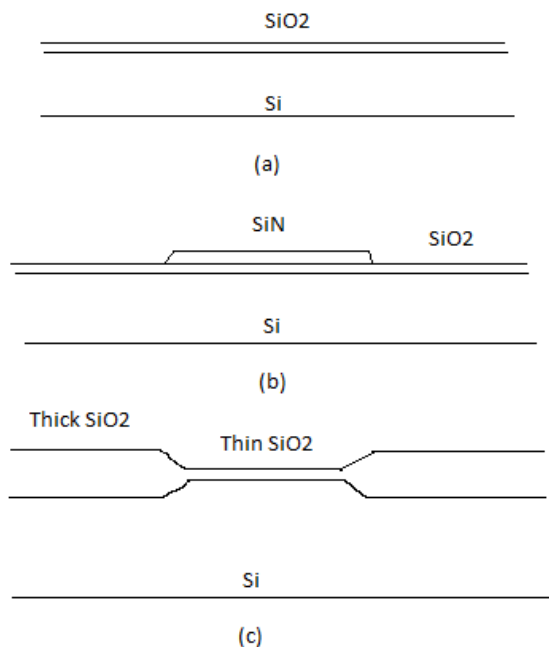


Fig. 7 Local Oxidation Process (a) Si sample prior to the deposition of nitride (b) After the deposition of the nitride (c) After oxidation and nitride removal.²

In this process, first a layer of Silicon Nitride (SiN) is deposited over the substrate, which is followed by its removal with a masking step from all areas where a thick layer of dioxide is to be grown. The layer of silicon nitride

serves as a barrier to the oxidation. So when the high-temperature oxidation step takes place, a thick oxide is grown in the regions where the nitride is absent. The process can be made much clear with the following figures.

F. Polysilicon Deposit

The modern technologies are making use of some layers of polycrystalline silicon that are deposited during the fabrication. Once the layers of the polycrystalline silicon are deposited, all the desired features are defined by employing the masking step, and can serve for various electronic applications. Another feature is that, the sheet resistance of the layers can be controlled by the impurity added. Polysilicon can be undoped or doped with elements such as As, P, or B to reduce the resistivity.

Polysilicon serves as:

- Gate electrode material in MOS devices
- Conducting materials for multilevel metallization
- Contact materials for devices with shallow junctions.

3. FABRICATION TECHNIQUES FOR MEMS

MEMS fabrication uses many of the same techniques that are used for fabrication of integrated circuits such as oxidation, diffusion, ion implantation, etc. along with highly specialized micromachining processes [11].

Some of the most widely used micromachining processes are discussed below.

A. Bulk Micromachining

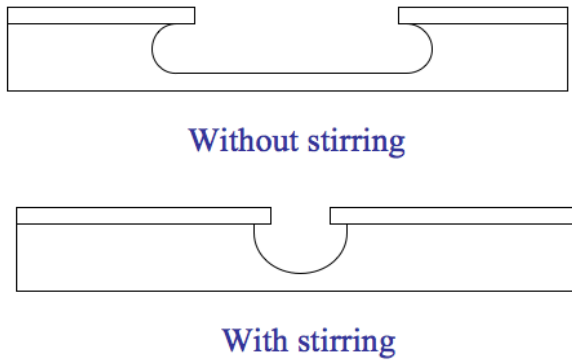
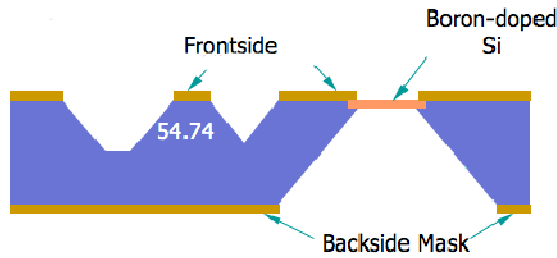
It is the oldest micromachining technique. In this technique selective removal of substrate material is done in order to produce miniaturized mechanical components. This technique can be used either by using chemical or physical means.

Mostly wet chemical etching is used because it provides high etch rate and selectivity. There are two types of wet chemical etching.

1) Isotropic wet etching

Isotropic wet etching doesn't depend on the crystallographic orientation of the substrate and the etching process is carried out in all directions. Practically, lateral etching is slower than normal etching without stirring. Thus, isotropic etching is done while vigorously stirring the etchant solution.

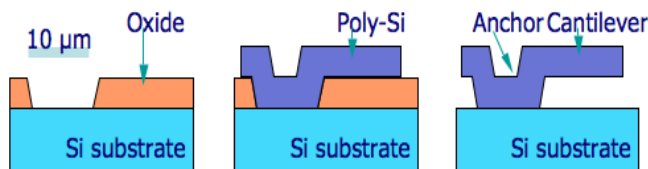
As can be seen in Figure 8, the lateral etching takes place when no stirring is done so as to prevent it vigorous stirring is done.

Fig. 8 Isotropic wet etching³Fig. 9 Anisotropic wet etching⁴

2) Anisotropic wet etching

Anisotropic wet etching involves dipping of substrate into a chemical solution where etch rate depends upon the crystallographic orientation of the substrate. In such process etching is done at a higher rate normally than laterally. This type of etching is direction specific as can be seen in Figure 9.

B) Surface Micromachining

Fig. 10 Surface Micromachining⁵

The fabrication involves certain steps. Firstly, the deposition of thin oxide film is done so as to act as a temporary mechanical layer upon which the actual devices layers are built. Followed by the deposition and patterning of device layer of material also known as structural layer, then the temporary layer is removed so as to release the micromechanical structure layer.

3) MEMS Pressure Sensors

Conventional micro-machined piezoresistive pressure sensors have a silicon substrate with a back etched membrane which supports four piezoresistors arranged in such a way that they coincide with the positions of maximum sensitivity for pressure-induced deflections of the diaphragm, as shown in Figure 11(a). A wet etching process is widely used to form the back-etched hollow from the backside of the pressure sensor wafer. To make a closed vacuum chamber in the pressure sensor, the pressure sensor wafer is bonded with a Pyrex glass wafer using an anodic bonding process in a vacuum environment. The back-etched hollow of the pressure sensor and the bonded Pyrex glass, therefore, form a closed vacuum chamber. In practice, the resistance changes of the piezoresistors vary as a function of membrane stress and piezoresistive coefficients of the piezoresistors of the pressure sensor in the longitudinal and transverse directions. The piezoresistive coefficients of piezoresistors vary as a function of both the temperature and doping concentration [12].

Figure 11(b) shows the resistors are electrically connected in a Wheatstone bridge layout. When pressure is applied to the silicon membrane of a pressure sensor, it deflects in the downward direction, causing a change in the resistance values of the four piezoresistors. This in turn induces a change in the output voltage of the Wheatstone bridge. The environment pressure can be derived using the output voltage change of the pressure sensor. The current study considers the package of a p-type silicon pressure sensor deposited in an n-type epi-layer with a specific thickness grown on a p-type silicon substrate. The square silicon membrane of the piezoresistive pressure sensor has an area of $576 \times 576 \mu\text{m}$ and a thickness of $20 \mu\text{m}$. It is designed for an absolute pressure range of $0-690 \times 10^3 \text{ Pa}$ with TCO of $0.2\% \text{ span}/^\circ\text{C}$ [13].

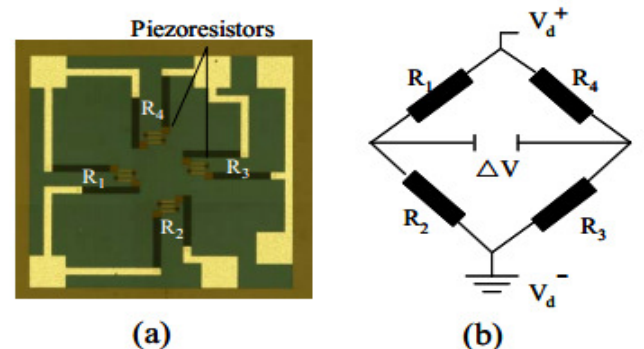


Fig. 11. Piezoresistive pressure sensor chip: (a) upper surface showing four piezoresistors; (b) Wheatstone-bridge layout of four piezoresistors.¹²

4. CONCLUSION

Integrated circuits are the future of electronics. Fabrication of integrated circuits can be done by various above mentioned processes. These days' photolithography and ion implantation techniques are mostly being used for the fabrication of integrated circuits. Micro-Electro-Mechanical System (MEMS) is an advancement of integrated circuits which helps in reduction of the size of mechanical components and electronic circuits. These MEMS such as piezoresistive pressure sensors can be fabricated by various micromachining processes mentioned above.

REFERENCES

- [1] D. Choudhary Roy, "Linear integrated circuits".
- [2] N. Maluf, "An introduction to microelectromechanical systems engineering".
- [3] Gary, Hurst, Lewis and Meyer, "Analysis and design of analog integrated circuits".
- [4] Mattias Torstensson, "Photolithography".
- [5] W.M. Moreau, "Semiconductor Lithography - Principles, Practices and Materials", Plenum Press, New York, USA, 1988.
- [6] C.Y. Chang and S.M. Sze, "ULSI Technology", McGraw-Hill, New York, USA, 1996.
- [7] M. Alonso and E.J. Finn, "Physics", Addison Wesley, Harlow, England, 1992.
- [8] Marc Walker, "An introduction to lithography".
- [9] Justinas Palisaitis, Remigijus Vasiliauskas, "Epitaxial growth of thin films", Linköping University, Sweden.
- [10] Nathan Cheung, "Ion Implantation", U.C. Berkeley.
- [11] Yilong Hao, "Silicon based MEMS fabrication techniques and standardization".
- [12] Smith, C.S., "Piezoresistance effect in germanium and silicon". Phys. Rev. 1954.
- [13] Lung-Tai Chen, Jin-Sheng Chang, Chung-Yi Hsu, and Wood-Hi Cheng, "Fabrication and performance of MEMS-based pressure sensor packages using patterned ultra-thick photoresists".

Comparison between the Implementation of data Compression using Huffman Coding and Shannon-Fano Coding through VHDL

Pooja Srivastava¹, Jyotsna Joshi², Abhishek³

^{1,2,3}Department of Electronics, SRMCEM,
¹pooja_enn@yahoo.com, ²jyotsna.devna@gmail.com, ³abverma1991@gmail.com

Abstract: The digital era embarked upon by the evolution of computers has revolutionized the mankind. Present system requirements are determined by its ability to store large volumes of data set. Meanwhile, the enhanced capacity of communication networks has opened new avenues for transfer of enormous data over communication links. In order to make an optimum use of the communication channel and other resources, data compression has to be carried out effectively. This compressed data, in turn, has to be decoded efficiently, which is the theme of this paper. It aims towards the comparative study of the implementations of a high speed Huffman decoding system and Shannon-Fano decoding system. This proposed model gives an insight about the speed of decoding operations in both the cases. The model is implemented using VHDL language, simulated on Xilinx ISE Version '6.1i' Model Sim Simulator.

Keywords: Huffman Decoder, Shannon-Fano decoder, Data compression.

1. INTRODUCTION

The major problem existing with the current compression and encryption methods is the large amount of processing time required by the computer to perform the tasks^[1]. This problem, however, can be dealt with the idea of adding a pseudo random shuffle into a data compression process. As a matter of fact, simultaneous data compression and encryption offers an effective remedy for the execution time problem in data security.

Huffman coding is an entropy encoding algorithm used for lossless data compression. It is a popular method for compressing data with variable-length codes^{[2] [3]}. The term refers to the use of a variable table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol. Also known as Static Huffman coding, Huffman's algorithm provided the first solution to the problem of constructing minimum-redundancy codes. Huffman coding method encodes the input data with shortest code word for highly occurring symbols, and long code word for less occurring symbols^[4]. This results in

higher compression of data set with the reduction in decoding time when compared to fixed length coding.

Whereas, in Shannon-Fano coding, the symbols are arranged in order from most probable to least probable, and then divided into two sets whose total probabilities are as close as possible to being equal. All symbols then have the first digits of their codes assigned; symbols in the first set receive "0" and symbols in the second set receive "1". As long as any sets with more than one member remain, the same process is repeated on those sets, to determine successive digits of their codes. When a set has been reduced to one symbol, of course, this means the symbol's code is complete and will not form the prefix of any other symbol's code.

The difference between the two algorithms is that the Shannon-Fano code tree is built from the top down, while the Huffman code tree is constructed from the bottom up.

2. PROPOSED METHODOLOGY

On the basis of the 11 8-bit input data applied using behavioural modelling approach in VHDL, we have designed encoder for both Huffman Coding & Shannon Fano Coding. From the encoded word we get the information about the occurrence of different inputs using which we will compare both the coding techniques on the basis of the comparison parameters. Fig.1 illustrates the proposed methodology.

1. Input: The input of our project is the eleven data words each consisting of 8 bits. All the inputs should not be identical (though few of the words may be repeated). This is done so that the more frequently occurring words are assigned shorter code & less frequently occurring words are assigned longer codes so that a suitable binary tree (which is not possible if all the data words are same) can be designed (in case of Huffman coding) using these codes.

2. Occurrence Calculator: The basic purpose of the occurrence calculator is to evaluate each of the eleven 8 bit

data & then calculate the frequency of each word, i.e., the number of times a word is repeated.

3. Sorting Signal: The sorting of all the data words is done in descending order on the basis of the frequency of each word, i.e., the number of times each data word is repeated. The larger the frequency of a data word, the larger is its probability of occurrence and the higher is its position in the sorting table.

4. Huffman/Shannon Fano Encoder: The encoder assigns an appropriate code-word to each eleven 8-bit data word. The more frequently occurring data words are assigned shorter codes & less frequently occurring words are assigned longer codes. This is done by implementing Huffman/Shannon Fano Coding in VHDL.

3. IMPLEMENTATION:

A. Huffman Algorithm:

The Huffman tree building algorithm can be described by the following five steps:

Step 1: Create an initial list of characters (sub trees) and frequencies.

Step 2: If the sub tree list contains only one tree item then you are done.

Step 3: Otherwise, remove from the list the two sub trees with the smallest frequencies and make them children of a new combined sub tree whose frequency is the sum of the two child frequencies.

Step 4: Add the new combined sub tree to the list, keeping the entire list sorted by frequency.

Step 5: Go back to step 2^[5].

Let us construct the Huffman code for the following set of messages: x1, x2, x3, x4,

x5 with the probabilities $p(x1) = \dots = p(x5) = 0.2$

1. x1 (p=0.2), x2 (p=0.2), x3 (p=0.2), x4 (p=0.2), x5 (p=0.2)
2. x4, x5 \rightarrow x45 (p=0.4) \Rightarrow x45, x1, x2, x3
3. x2, x3 \rightarrow x23 (p=0.4) \Rightarrow x45, x23, x1
4. x1, x23 \rightarrow x123 (p=0.6) \Rightarrow x123, x45
5. x123, x45 \rightarrow x12345 (p=1)

Fig.2 below shows the generated Huffman tree corresponding to the parameters in the above example.

B. Shannon- Fano Algorithm

Step 1: For a given list of symbols, develop a corresponding list of probabilities or frequency counts so that each symbol's relative frequency of occurrence is known.

Step 2: Sort the lists of symbols according to frequency, with the most frequently occurring symbols at the left and the least common at the right.

Step 3: Divide the list into two parts, with the total frequency counts of the left half being as close to the total of the right as possible.

Step 4: The left half of the list is assigned the binary digit 0, and the right half is assigned the digit 1. This means that the codes for the symbols in the first half will all start with 0, and the codes in the second half will all start with 1.

Step 5: Recursively apply the steps 3 and 4 to each of the two halves, sub-dividing groups and adding bits to the codes until each symbol has become a corresponding code leaf on the tree^[6].

Tables 1 & 2 illustrate this approach.

C. Comparison Parameters:

The comparison between Huffman Coding & Shannon Fano Coding is done on the basis of 3 parameters:

1. *Average Code-word Length (L):* It is defined as the average number of bits per message.
2. *Source Entropy (H):* It is defined as the average information per message, as given by Eqn.(1).
3. *Code Efficiency (η):* It is defined as the ratio of the average information per message (H) and the average code word length (L), as given by Eqn.(2).

According to the above parameters, the results of the two algorithms can be summarised as below.

Huffman:

Encoding vectors: x1 \rightarrow (00); x2 \rightarrow (010); x3 \rightarrow (011); x4 \rightarrow (10); x5 \rightarrow (11)

1. Entropy: $H(X) = -5(0.2 \log 0.2) = 2.32$

2. Average length of the encoding vector:
 $L = 3(0.2 * 2) + 2(0.2 * 3) = 2.4$

3. The Huffman code gives $(2.32/2.4)100\% = 97\%$ efficiency.

Shannon-Fano:

Table 1. Various symbols & corresponding occurrence & their probabilities

Symbol	A	B	C	D	E
Count	15	7	6	6	5
Probability	.38461538	0.17948718	.15384615	.15384615	.12820513

Table 2: Shannon-Fano codes generated after applying algorithms

Symbols	A	B	C	D	E
Code	0	100	101	110	111

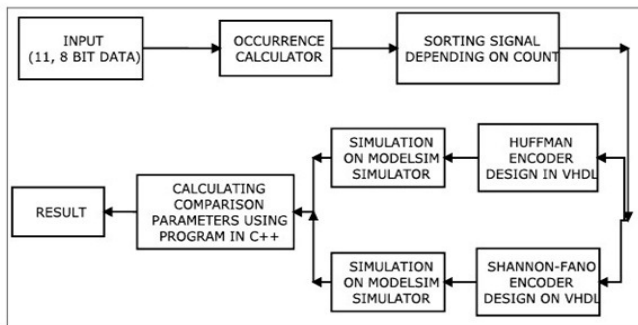


Fig.1 Block diagram of proposed methodology

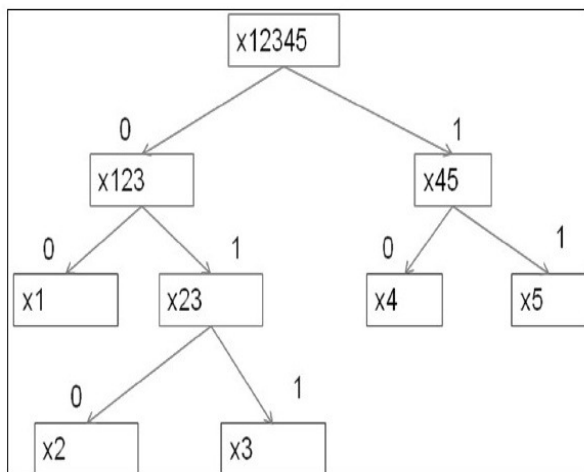


Fig.2 Huffman tree

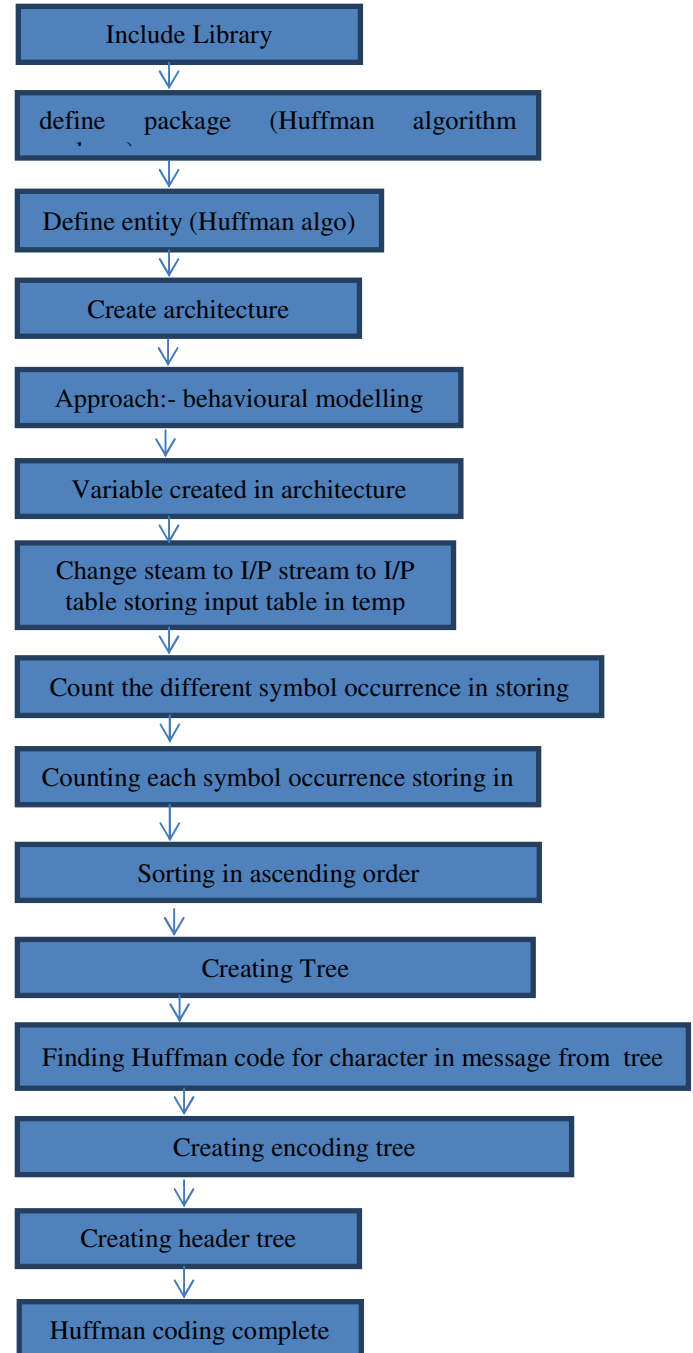
A comparative study between the flowcharts of the two algorithms is depicted in the next page.

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = \sum_{i=1}^n p(x_i) \log_2 \frac{1}{p(x_i)} = - \sum_{i=1}^n p(x_i) \log_2 p(x_i), \quad (\text{Eqn.1})$$

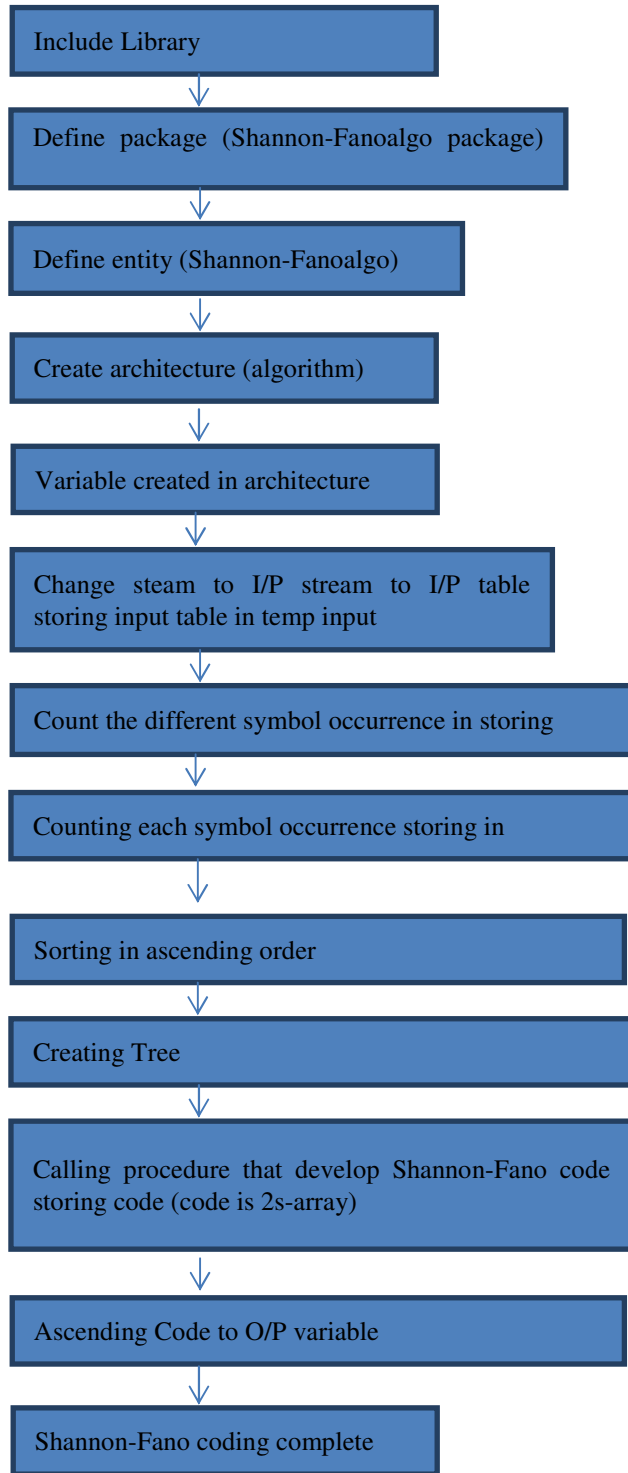
$$\eta = (H/L) * 100\% (\text{Eqn. 2})$$

D. FLOWCHART:

Huffman coding



Shannon-Fano coding



4. RESULT

Here results are simulated for the input
 —01100110011000011001010110011101101000011001100110
 01000011001010110
 01100110011001100110.
 That is 11, 8-bit data, a total of 88 bits.

5. RESULT IN FORM OF WAVEFORMS:

The following simulations^{[7], [8], [9], [10]} of both Huffman and Shannon-Fano coding are illustrated by taking a sequence of 5 inputs. For Shannon-Fano coding, the occurrence and bits needed for the 5 input sequences are 5&2, 2&2, 2&2, 1&3 and 1&3 respectively. The average code-word length, the entropy and the code efficiency are 2.181818 bits/message, 2.040373 bits/message and 93.517105% respectively. Similarly for Huffman coding the occurrence and bits needed for the 5 input sequences are 5&1, 2&2, 2&3, 1&4, 1&4 respectively. The average code-word length, the entropy and the code efficiency are 2.090909 bits/message, 2.040373 bits/message and 97.583061% respectively.

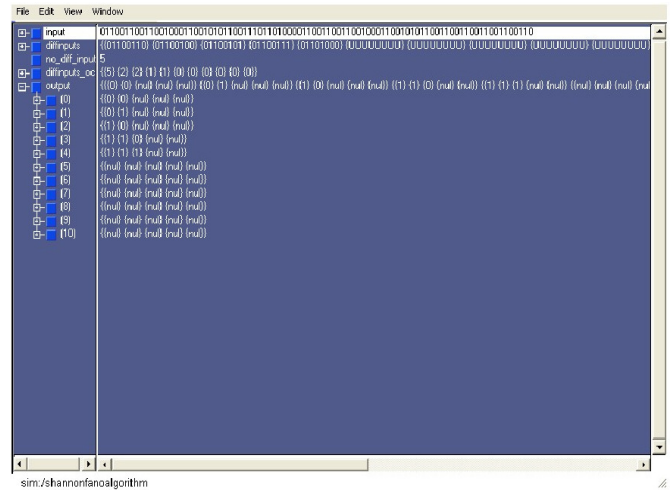




Fig. 5. Huffman Coding output signal



Fig. 6. Huffman Coding output wave form

The high speed Huffman and Shannon-Fano coding are compared on the basis of which the following points are noted:

Huffman Coding tends to offer significantly better compression for small alphabet sizes as compared to the other coding technique.

Huffman coding offers a high speed as compared to Shannon-Fano coding.

Lossless data compression techniques are used in communication system to transfer data efficiently over a channel, where prime requirement is of saving bandwidth.

6. ADVANTAGES:

1. Huffman coding is a statistical technique which attempts to reduce the amount of bits required to represent a string of symbols.
2. Huffman is widely used in all the mainstream compression formats that you might encounter - from GZIP, PKZIP (win zip etc.) and BZIP2, to image formats such as JPEG and PNG.
3. Image Compression.
4. Latest application of data compression technique is in Windows 8.
5. Fingerprint Patterns Recognition.

7. CONCLUSION AND FUTURE SCOPE:

From the results obtained we can see that Huffman Coding is far more superior, simple and faster to implement. For a given input, the amount of data compressed in Huffman Coding is more than that of Shannon Fano Coding.

Comparing Huffman algorithm to others, such as Shannon Fano coding, not many new features arise, although it uses a method capable of being applied to characters of any size. It does depend upon a pretended key that, if hidden and known to both the sender and the receiver, can be used for elementary encryption, though perhaps at a loss of some compression.

It is interesting to note that the Huffman coding algorithm, originally developed for the efficient transmission of data, also has a wide variety of applications outside the sphere of data compression. These include construction of optimal search trees [Zimmerman 1959; Hu and Tucker 1971; Itai 1976], list merging [Brent and Kung 1978], and generating optimal evaluation trees in the compilation of expressions [Parker 1980]. Additional applications involve search for jumps in a monotone function of a single variable, sources of pollution along a river, and leaks in a pipeline [Glassey and Karp 1976]. The fact that this elegant combinatorial algorithm has influenced so many diverse areas underscores its importance. Since the speed of data transmission is increased, many of the organizations and IT companies are using it for online data transmission.

REFERENCES:

- [1] K. Ashok Babu and V. Satish Kumar, "Implementation of Data Compression using Huffman Coding", 2010 International Conference on Methods and Models in Computer Science (ICM2CS-2010).

- [2] R.C. Gonzalez and R.E. Woods, "Digital Image Processing" Second edition.
- [3] John Kennedy, "Huffman Coding", Mathematics Department Santa Monica College 1900 Pico Blvd. Santa Monica, CA: 90405.
- [4] LaurentiuAcasandrei and Marius NEAG, "A Fast Parallel Huffman Decoder for FPGA Implementation", ACTA TECHNICA NAPOCENSIS (Electronics and Telecommunications), Volume 49, Number 1, 2008.
- [5] S. T. Klein and Y. Wiseman, "Parallel Huffman Decoding with Applications to JPEG Files", The Computer Journal, 46(5), copyright British Computer Society, 2003.
- [6] D. Coelho, the VHDL Handbook, Boston: Kluwer Academic, 1988.
- [7] J. Bhaskar, A VHDL Synthesis Primer, Allentown, PA: Star Galaxy Publishing, 1995.
- [8] Z. Navabi, "VHDL-Modular Design and Synthesis of cores and Systems", TMH – 3rd Edition.
- [9] C. H. Roth, "Digital System Design using VHDL", PWS Publishing.
- [10] R.D.M. Hunter, T. T. Johnson, "Introduction to VHDL" Springer Publication, 2010.

Design and Performance Analysis of Sound Level Meter

Sushil Kumar

*School of Information and Communication Technology
Gautam Buddha University
Greater Noida, Uttar Pradesh-201 310, India
sushilkumar0108@gmail.com*

Abstract: In this paper a root mean square (RMS) circuit is used in the sound level meter as an exponential average taking the AC signals from a microphone, converting it to DC by RMS. The RMS must have a time constant which is given in the terms of time weighting. The two standardized time weightings that will be investigated for the project are S for slow (1s) and F for fast (50ms). The output of an RMS is a linear voltage and is passed through a logarithmic circuit to give an output in dB.

The root mean square is a measure of the average power of a signal and uses mainly the voltage and current instead of powers.

Index Terms: RMS, DC, linear voltage, logarithmic circuit.

1. INTRODUCTION

The RMS is the root of the mean of the square. It is obtained by calculating the square of the voltage waveform to obtain a representing one of the change of power with time. This power is averaged and the square root of this is the RMS. A sound level meter is an instrument which measures the level of sound pressure. It is commonly used in noise pollution studies for the quantification of different kinds of noise especially aircraft noise, industrial and environmental [1]. Although, the reading from a sound level meter does not correlate well to loudness perceived by humans which is better measured by a loudness meter.

Sound level meters are divided into two classes. Sound level meters of the two classes have same functionality but different tolerances for error. Class one instruments have a wider frequency range and a tighter tolerance than a lower cost. This applies to both the sound level meter itself as well as the associated calibrator. The IEC 61672-1 specifies three kinds of sound measuring instruments. They are actually conventional, integrating-averaging and integrating sound level meter [2].

The standard sound level meter can be called an exponentially averaging sound level meter as the AC signal from the microphone is converted to DC by a root mean square (RMS) circuit and thus it must have a time constant

of integration, now a days referred to as the time weighting. Three of these time weightings have been internationally standardized S originally called Slow, F (125 ms) originally called Fast and I (35 ms) originally called Impulse. Their names were changed in the 1980s to be the same in any language. I time-weighting is no longer in the body of the standard because it has little real correlation with the impulsive character of noise events [3].

The output of the RMS circuit is linear in voltage and is passed through a logarithmic circuit to give a readout linear in decibels (dB). This is 20 times the base 10 logarithm of the ratio of a given root mean square sound pressure to the reference sound pressure [4]. Root mean square sound pressure being obtained with a standard frequency weighting and standard time weighting. The reference pressure is set by international agreement to be 20 micropascals for airborne sound [5].

An exponentially averaging sound level meter which gives a snapshot of the current noise level, is of limited usage for hearing damage risk measurements; an integrating averaging meter is usually mandated [6]. An integrating meter simply integrates or in other words sums the frequency weighted noise to give sound exposure and the metric used is pressure squared times time [7].

Section II deals with the system design of the sound level meter i.e; the circuit diagram of the same in which part A includes the Dynamic Range, part B includes Linearity and part C talks about the Linearity Tolerance Testing. Eventually, part III deals with the conclusion and the detailed analysis of the results obtained systematically throughout the paper.

2. CIRCUIT DIAGRAM

The op amp based rectifier provides precise AC to DC conversion. Providing accuracy even with millivolt AC signals. The signals are first rectified and filtered to get a smooth DC output. For random signals the desired RMS is

required. The circuit provides a precision absolute value circuit connected to a quadrant multiplier.

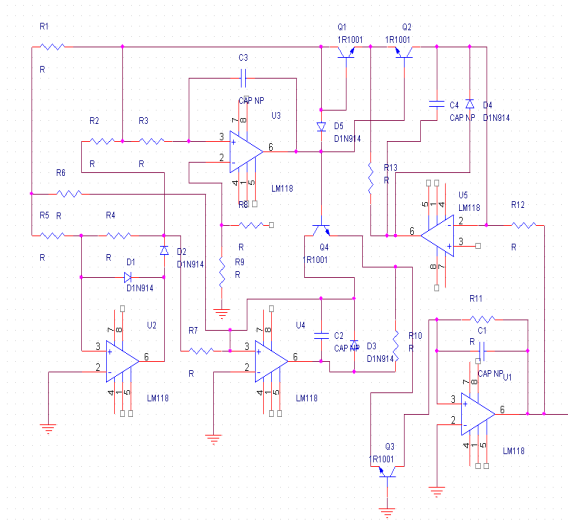


Fig. 1. RMS Detector

The bottom left amplifier provides the absolute value amplifier and provides a positive input current to the middle two amplifiers separate from the signal current. If the signal is positive the output of the absolute value stops the signal flowing through the connecting resistors. The amplifiers (not absolute value amplifier) form a log multiplier/divider.

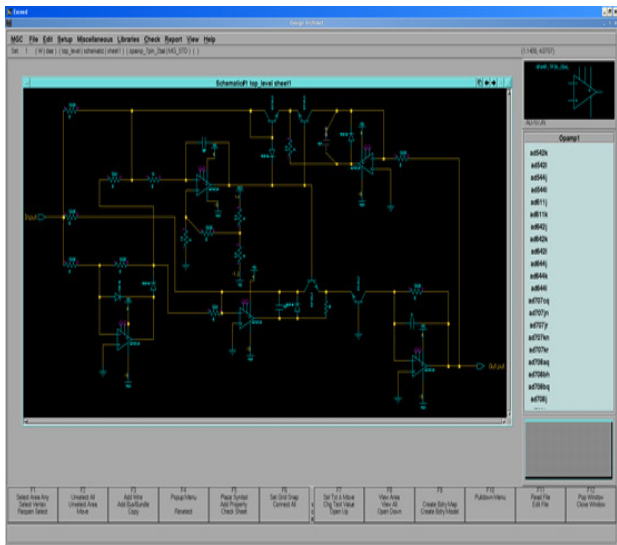


Fig. 2. Circuit in Mentor

Initial developments of the circuit show up several errors, with mentor stating connection problems. After initially not been able to see any problems the circuit needed to be rebuilt and this seemed to eradicate any issues. The first tests on the circuit showed inaccuracies. This was down to the tolerances

of components and suitability for the application. Changing the op amps from LM218 to AD707 and using BC109-c transistors gave it a good performance for an aim of 40dB range at grade 1.

A. Dynamic Range

To look at the dynamic range of the RMS detector testing has been done across the voltage range at 10dB increase to cover the 40dB requirement and with each increase the frequency will be increased to look at how it changes across the range of what it can receive. The primary area for this will be from 0.63V - 6.3V. The first test was done at 31.5Hz with 6.3mV. The following shows the output: (all other tests will not have outputs shown)

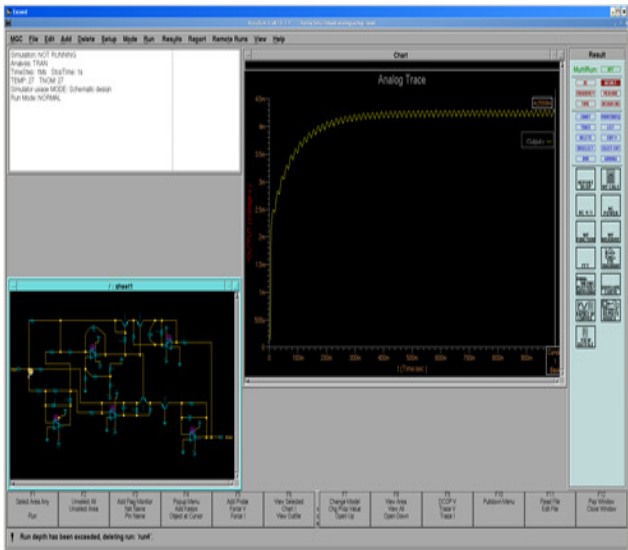


Fig. 3. Example Output for Tables Below

The first testing was done at 31.5Hz across the range of 40dB this will show what type the RMS falls into across the dynamic range.

TABLE 1. Dynamic Testing For 31.5 Hz

31.5Hz	Input RMS	Output	dB	Input dB	Output dB
	4.4541mV	4.2559mV	0.395	-47.0248	-47.42
	44.541mV	44.094mV	0.07	-27.0248	-27.112
	445.41mV	442.71mV	0.0528	-7.0248	-7.0776
	4.4541V	4.4391V	0.0293	12.9751	12.9458

The results show the difference between the primary indicator range and secondary indicator range with primary giving much better results. All of the results from the testing at this frequency proves to be type 1 which was specified before the development.

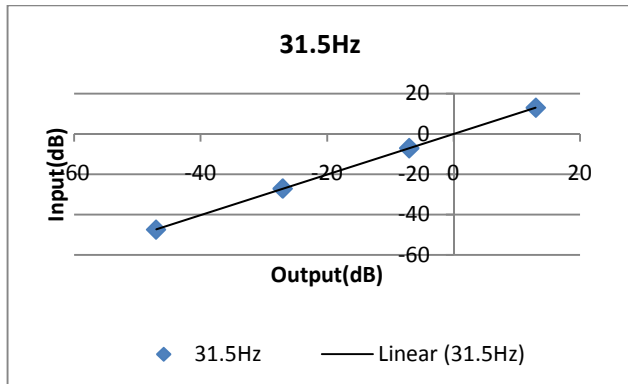


Fig. 4.31.5 Hz Input dB against Output dB

Figure 1.4 shows the linearity across the range of voltages at 31.5Hz. The same tests were done at 1KHz and 8KHz and the results are shown below.

Table II: Dynamic Results for 1 KHz

1KHz	Input RMS	Output	dB	Input dB	Output dB
	4.4541mV	3.7608mV	1.469	-47.0248	-48.4944
	44.541mV	43.263mV	0.252	-27.0248	-27.2776
	445.41mV	442.12mV	0.0643	-7.0248	-7.0892
	4.4541V	4.4441V	0.0195	12.9751	12.9556

The result from the testing at 1KHz starts to show how the lower input voltages start to fall away and become out of the range required other voltages soon become in range as it increases. This could be due to the turning on requirements of some components with such small voltages it need to be more ideal for the best results.

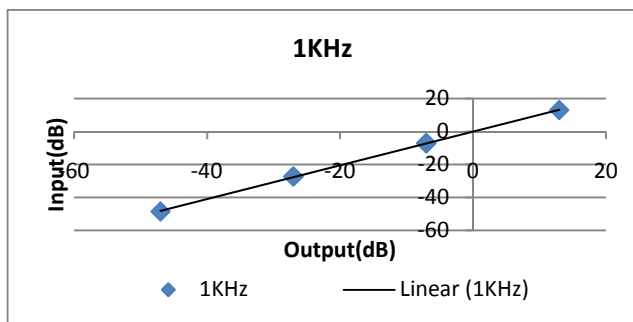


Fig. 5.1 KHz Input dB Vs Output dB

The plot in figure 1.5 show the system is still linear but at the end of the trendline at the lower voltage the linearity starts to fall away. This is an expected outcome and will not cause any problems as the increase in the frequency with the small voltages makes the signals become less visible.

Table III: 8 KHz Dynamic Range Testing

8KHz	Input RMS	Output	dB	Input dB	Output dB
	4.4541mV	2.6713mV	4.4407	-47.0248	-51.4596
	44.541mV	38.591mV	1.2455	-27.0248	-28.2703
	445.41mV	435.24mV	0.2006	-7.0248	-7.2254
	4.4541V	4.3549V	0.1956	12.9751	12.7795

The outputs showed in the table III shows once again that the smaller voltage at higher frequency completely falls off compared to the higher voltages. The higher the frequency the results become less accurate, this would require more tweaking on the circuit to correct any issues. Component values of the capacitors that enable the fast and slow timing will adjust the resulting outcomes.

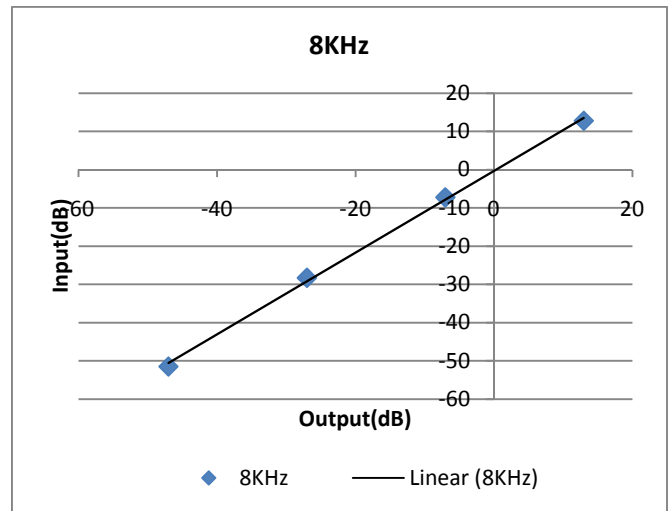


Fig. 6. 8 KHz Input dB against Output dB

The figure above showing the trendline is becoming less accurate although a linear response exists it is starting to lose accuracy at the furthest point from the primary indicator. Most of the test show that it is within the range required for type 1 other than the lower voltage at high frequency but this is the worst case scenario, while other values are well under the designated levels.

B. Linearity

Using a DC sweep of the circuit to look at the linearity gave good results and the following figure shows the resulting output. The marker on the output shows that at 385mV on the input the output was 383.91mV giving a difference of 1.9mV which is good.

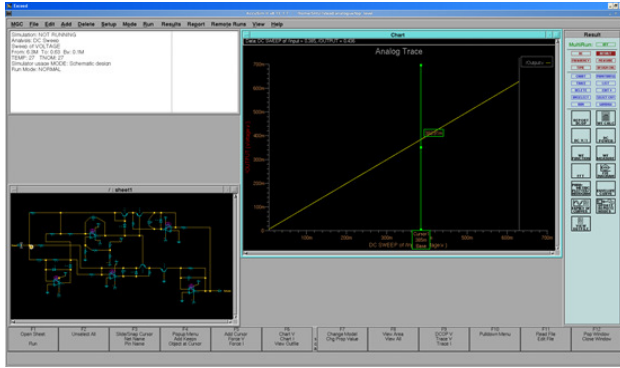


Fig. 4. DC sweep from 6.3 mV to 630 mV

Converting the axis to a logarithmic scale gives the following curve showing at what voltage the input will give the best outputs. From the scale it can be seen that it does not response well until about 100 mv.

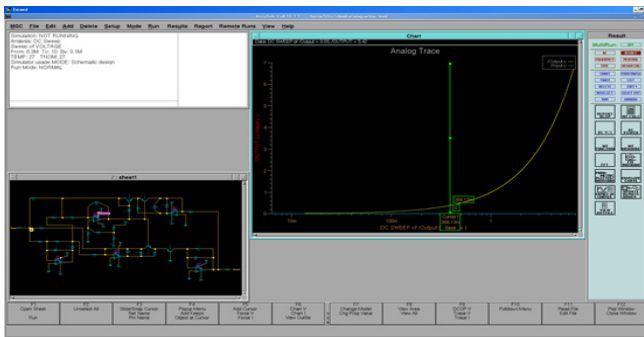


Fig. 5. Logarithmic DC sweep

The following figure show how making adjustments to the potential divider on the middle amplifier changes the linearity when plotted against the input voltages. This is adjusted to give the smallest difference to the input voltage.



Fig. 6. Linear output with changes

C. Linearity Tolerance Testing

To compare the type 1 tolerance against the developed circuit and ranges outputs have been generated to look at the differences in input to output. With an input force of 1V magnitude and setup from 0 – 1s looking at 31.5Hz, 1kHz, and 8kHz and comparing the decibel difference. All of the following plots show the output generated.



Fig. 7. Linearity Tolerance Testing

The input of 1V magnitude will give an RMS of $\sqrt{2} \times 1 = 0.707$ so to find the different in dB the following equation has been used with the output maximum after 1s been 0.70223.

$$20 \times \log\left(\frac{0.707}{0.70223}\right) = 0.058 \text{ dB} \quad (1)$$

For type 1 to be achieved this need to be less than 1.1dB in the fast setting and this has easily been achieved.

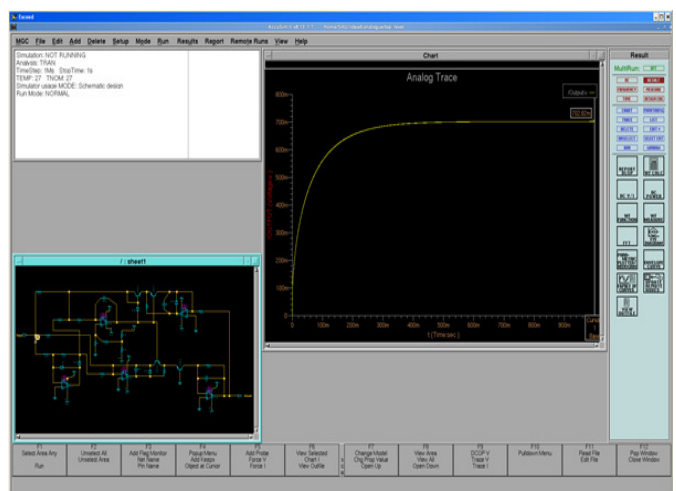


Fig. 7. 1V at 1 KHz (f)

Again using the same method as above to cover the range of accepted frequencies the figure above showing the output at 1KHz. The output at 1KHz is 0.70282 this gives the following difference.

$$20 \log \left(\frac{0.707}{0.70282} \right) = 0.051 \text{ dB} \quad (2)$$

This is also within the 1.1dB requirement so is still type 1. The same was done on at 8KHz and this gave an output of 0.69592.

$$20 \log \left(\frac{0.707}{0.69592} \right) = 0.137 \text{ dB} \quad (3)$$

To conclude these tests all of the result show that the circuits output at a range of frequencies give the required level of accuracy for type 1 or even type 0.

3. RESULT

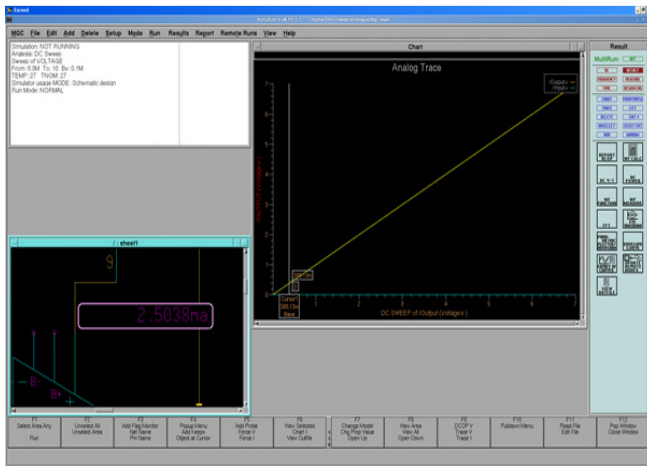


Fig. 8. Current drawn of the circuit at 2.5038 mA

Based on all testing the overall grade of the RMS it would be grade 1. Based on all of the test completed a lot more work could have gone into the tweaking the circuit to type 0, minor adjustment were required to take it to the next grade but due to earlier software problem it hasn't been completed.

The following figure 1.8 shows the current is drawn of the circuit at 2.5038mA. When integrated with the other parts from the sound level meter it will be a small amount of current drawn from the battery giving a longer battery life.

REFERENCES

- [1] K. Iniewski, Electronics for Radiation Detection. Boca Raton, FL: CRC Press, Jul. 2010.
- [2] W. M. C. Sansen, Analog Design Essentials. New York: Springer, 2006, p. 133.
- [3] J. Savoj and B. Razavi, "A 10-Gb/s CMOS clock and data recovery circuit with a half-rate linear phase detector," IEEE J. Solid-State Circuits, vol. 36, no. 5, pp. 761–767, May 2001.
- [4] B. Razavi, Design of Integrated Circuits for Optical Communications, 1st Ed. New York: McGraw-Hill, 2003, pp. 318–322.
- [5] John Sartori and Rakesh Kumar (2006), "Overscaling-friendly Timing Speculation architectures", IEEE J. Solid State Circuit, Vol. 41, PP. 792-804.
- [6] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw, T. Austin, T. Mudge, and K. Flautnerm 2006, "A self-tuning DVS processor using delay-error detection and correction," IEEE J. Solid-State Circuits, pp. 792–804.
- [7] Jan M Rabaey, "Digital Integrated Circuits, A Design Perspective", Prentice Hall, 2003.

Microcontroller Based Data Analyzer through Wireless RF

Akshay Nangia¹, Madhurima Basak², Nippun Bahl³

^{1,2,3}Amity Institute of Telecom Technology & Management, Amity University, U.P., Sector-125, Noida-201303

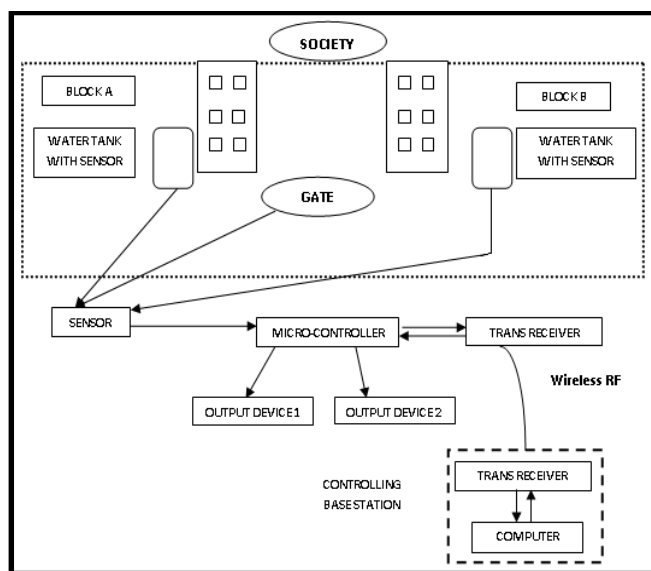
¹akshaynangia1234@gmail.com, ²madhurima.1991@yahoo.in

Abstract: Nowadays, Residential societies / Localities face the major problems of water level monitoring for its tanks, security, soil water sensing and other issues. Therefore, there is an immediate requirement of an automated system or a device which controls the daily requirements of a society as mentioned. This paper represents the prototype design of microcontroller based data analyzer through RF ID which will allow the pumping of water in the appropriate tank and will ensure security by detecting intrusion.

Keywords: Water level monitor, metal sensor, microcontroller.

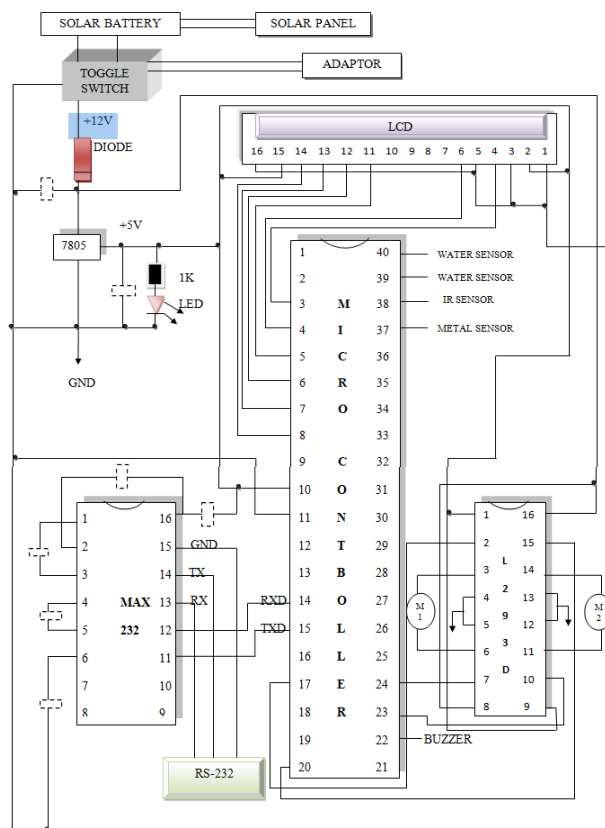
1. INTRODUCTION

In the present era, the systems used for daily requirements like water level monitoring and security in the locality/society are manually operated. The water level in the tanks cannot be monitored such that even if the tank is full, the water pump remains on and hence overflow may occur, thereby creating water wastage. Also a major issue in localities is that security of the society cannot be monitored. [1] There is a dearth of an intelligence system within the society or at the main gate that would detect any intrusion of visitors carrying unintended metal objects. This dearth leads to the designing of such a system.



There is a need of an automatic system which replaces the manually operated systems and help in monitoring the water level in the tanks and help in the maintenance of security of the society. [3] With the help of microcontroller and sensors (metal sensor, water sensor, IR sensor) the system can be implemented. The water sensor monitors the water level in the tanks and notifying the microcontroller [3] when to switch on and switch off the pump thereby preventing from overflow and water wastage. The IR sensor at gate helps to monitor the number of visitors in the society and displays it on the LCD screen. The metal sensor at the gate detects any metal intrusion and helps to maintain security of the society. Also, the use of solar energy with the help of solar panel helps to save the non-renewable resource of energy.

2. CIRCUIT DIAGRAM AND DESCRIPTION



The main components for system design are mentioned as follows:

- ATMEGA-16 microcontroller
- RF modules(CC2500)
- IC - 7805
- IC – L293D
- IC- MAX 232
- Sensors (IR, metal, water)
- RS- 232 connector
- LCD [5]

ATMEGA-16 microcontroller

The ATMEGA-16 microcontroller [4] is an 8-bit high performance microcontroller based on enhanced RISC (Reduced Instruction Set Computing) architecture. It has 16 KB programmable flash memory, static RAM of 1 KB and EEPROM of 512 Bytes. It is a 40 pin microcontroller. There are 32 I/O lines which are divided into four 8-bit ports designated as PORT A, PORT B, PORT C and PORT D.

IC- 7805

The IC 7805 is a series of three-terminal positive regulator with several fixed output voltages, making them useful in a wide range of applications. If adequate heat sinking is provided, they can deliver over 1A output current. These devices can be used with external components to obtain adjustable voltages and currents. Used here as a 12V- 5V converter.

RF ID module (CC2500)

Operating in the 2.4 GHz frequency band, the device includes several useful digital features like full packet handling, FIFO buffers, clear channel assessment, wake-on-radio and more. It works on a voltage of 5V (DC) only and current < 50mA. Also the operating frequency is 125 kHz.

RS 232 connector

RS-232 is the traditional name for a series of standards for serial binary single-ended data and control signals connecting between a Data Terminal Equipment and a data Circuit terminating Equipment. It is commonly used in computer serial ports. The standard allows as many as 20 signals to be defined, but gives complete freedom to the user. Three wires are sufficient: send data, receive data, and signal ground. The remaining lines can be hardwired on or off permanently. The signal transmission is bipolar,

requiring two voltages, from 5 to 25 volts, of opposite polarity.

IC- L293D

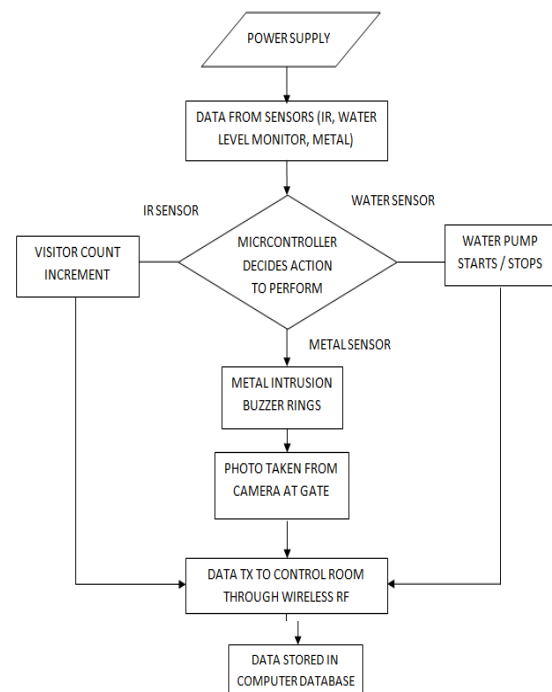
This device is a monolithic integrated high voltage, high current four channel driver designed to accept standard DTL or TTL logic levels and drive inductive loads (such as relays solenoids, DC and stepping motors) and switching power transistors. This device is suitable for use in switching applications at frequencies up to 5 kHz.

Sensors

The sensors used here are water level sensor, IR sensor for visitor count, metal sensor for metal intrusion detection

3. WORKING

The microcontroller circuit works with the power supplied from the 12V to 5V converter i.e. IC-7805. The input to the circuit is from a toggle switch which switches between solar energy and the power from a 12V adapter.



The solar panel is attached to the toggle switch which captures the solar energy in the day time or at the time of presence of solar energy. At the time when there is no solar energy or during dark, the circuit will switch to the main power supply from the adapter. All the devices except the pump used in this system work on the 5V voltage provided by the IC-7805. The power supply input to the microcontroller helps in conserving electricity and thereby making this an eco-friendly project.

The water sensor is present at the water tanks monitoring the water level. The sensor sends a pulse whenever the water level falls below the set threshold level.

The IR sensor [6] is present at the society gate to monitor the number of visitors in the society. Whenever someone enters through the gate the sensor sends a pulse to the microcontroller.

The Metal sensor is also present at the gate. If a person enters the society gate with a metal object, the sensor sends a pulse to the microcontroller. The analog signals from the sensors are fed to the microcontroller via pins of PORTA. The signals from microcontroller are transmitted serially through wireless RF using MAX-232.

The MAX-232 IC is connected to the RxD and TxD pins of the microcontroller (pin 14 and 15) for connection between the RS-232 connector and the microcontroller.

The microcontroller responds to each of these sensors by performing the appropriate actions. When the metal sensor sends a pulse, the microcontroller sends a pulse to the buzzer at the gate indicating a security breach has occurred and also to the security camera present at the gate which clicks a photograph of the person carrying the metal object and stores it in the database.

When the water level sensor sends a pulse to the microcontroller then it sends a pulse back to the water pump indicating it to shut down or start as per the signal sent by it earlier (water below threshold level or water above threshold level). All this record is maintained in a database stored in the computer at the control room which is done via the RF transceiver modules.

When the IR sensor sends the signal to the microcontroller then it sends a pulse and monitors the visitor count and displays the count on the LCD screen.

The computer present at the control room receives all the data from the microcontroller through wireless RF and it uses MATLAB image processing tools to capture the image and to convert all the data into a MS-EXCEL worksheet for storage.

4. CONCLUSION

The prototype was successfully developed and met the objectives. The system can automatically manage the major problems of water level monitoring for its tanks, security, soil water sensing and other issue. The requirement of an automated system or a device which controls the daily requirements of a society as mentioned is rising. This prototype design of microcontroller based data analyzer through RF ID which will allow the pumping of water in the appropriate tank and will ensure security by detecting intrusion can further be improved. In addition, the system is very practical when alternative and renewable resource of energy is used i.e. solar energy.

REFERENCES

- [1] Wireless RF data communications using 60 GHz antennas in Multi-Core systems by Ho-HsinYeh, Department of Electronics & Computer Engineering, Univ. of Arizona, Tucson, AZ, USA
- [2] Fang Meier, D.D., Garrote, D.J., Mansion, F. and S.H. Human. 1990. Automated Irrigation Systems Using Plant and Soil Sensors. In: Visions of the future. ASAE Publication 04-90, American Society of Agricultural Engineers, St. Joseph, Michigan.
- [3] Muhammad Ali Mazidi and Janice GillispieMazidi, "The 8051 microcontroller and embedded systems", Pearson education Ltd., India, 2007.
- [4] National semiconductor corporation, ADC 0809 data sheet, 8-bit Microprocessor compatible A/D converters with 8-channel multiplexer, national Semiconductor data book, October 2002 updates.
- [5] Introduction to LCD programming tutorial by Craig Steiner Copyright 1997 - 2005 by Vault information services LLC.
- [6] H.S.Kalsi, "Electronics and instrumentation", Tata McGraw-Hill Ltd., New Delhi, 1999

Design and Performance of Eleven Stages CMOS Ring Oscillator in 45 nm Technology

Sushil Kumar

*School of Information and Communication Technology
Gautam Buddha University
Greater Noida, Uttar Pradesh-201 310, India
sushilkumar0108@gmail.com*

Abstract: The paper deals with the design and performance analysis of a ring oscillator using CMOS 45 nm technology process on Cadence tool. The design of optimal analog and mixed signal very large scale integrated circuit is a challenging task for the integrated circuit designer. There are a number of challenges ahead while designing the CMOS ring oscillators which are delay, noise, total harmonic distortion (THD) and glitches. Being CMOS is the technology of many applications, CMOS oscillators with low timing jitter and phase noise are highly desired. Earlier, researchers were unable to reduce the phase noise in ring oscillators substantially with nine stages. The phase noise has successfully been reduced using Cadence virtuoso environment to -6.4kdBc/Hz at 2 GHz center frequency of oscillation.

Keywords: CMOS 45 nm Technology, Analog and Mixed Signal, CMOS Ring Oscillator, Integrated Circuit, Total Harmonic Distortion, Phase Noise.

1. INTRODUCTION

A ring oscillator is one of the most widely manufactured integrated circuit. Foundries use ring oscillators on every semiconductor wafer to monitor the gate delay and speed power product of fabricated MOS inverters. The automated measurements of oscillation frequency determine which wafers are acceptable and which fall outside an acceptable window and must be discarded. The ring oscillators have occupied this role since the earliest days of metal oxide semiconductor integrated circuit technology because they are easy to build always oscillate and are readily measured. The ring oscillator is a closed loop comprising an odd number of identical inverters which forms an unstable negative feedback circuit. Its period of oscillation is twice the sum of the gate delays in the ring because these oscillators are so well known to digital and analog circuit designers alike they have found use beyond the monitoring of the semiconductor process in communications circuits and clock generation [1]. A voltage controlled ring CMOS inverter based oscillator is first used for clock recovery in an ethernet controller [2]. Ever since the ring oscillator has become a widely used component in the communication technology. Today, almost all ring oscillators use differential delay stages because of their greater immunity to supply

disturbances[3]. Still, the ring oscillator is still the most widely fabricated of all oscillators. Firstly, compared to alternatives such as the LC resonator based oscillator, the ring oscillator is exceptionally compact. Secondly, it can oscillate at very high frequencies i.e; at very short periods limited only by the sum of a few gate delays. The maximum oscillation frequency is always much higher than LC phase shift oscillators although not as fast as LC oscillators that can tune a transistor to oscillate at its maximum frequency. Thirdly, since the ring oscillator is tuned in frequency by a current, its tuning range can span upto orders of magnitude. Only the relaxation oscillator, which is also tuned by a current, offers a similar tuning range. The right number of stages are chosen to oscillate at the desired frequency based on delay data given by the foundry. This is refined by a series of simulations. In today's time, the time jitter or phase noise in the oscillation can also be simulated. However, the analysis for gate delay becomes increasingly non intuitive as it gets more accurate, and the latest editions of textbooks on VLSI designs hold that is better to use the simplest possible analysis for a first order estimate of gate delay and then refine it with simulation [4]. The voltage controlled oscillator is the most important part of the Phase Locked Loop and generates the required frequency. An accurate estimate of the delay through a CMOS inverter loaded by the capacitance of a similar inverter is important not only for our purposes here, but is at the very center of the enterprise of VLSI design. The delay through a gate with fanout of one sets the absolute upper limit on clock frequency for a logic block.

The first paper estimates jitter caused by FET noise in CMOS differential ring oscillators cast the problem correctly in the time domain, by finding fluctuations in the instants when the output ramp in a delay element crosses the toggle point [5]. This is similar to the analysis used to find jitter and phase noise depends on voltage gain of the delay circuit. The next paper applies the concept of the impulse sensitivity function to the waveforms of a ring oscillator and from relation between the impulse sensitivity function and phase noise, deduces an approximate expression for phase noise [6]. The latest analysis of the ring oscillator on phase noise

explores details of the noise processes at the toggle point of the delay element jitter [7].

2. SYSTEM DESIGN

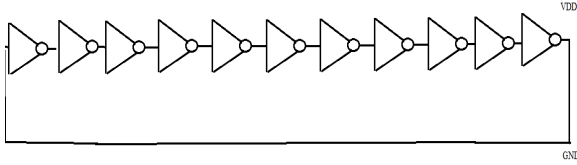


Fig. 1. Proposed block diagram of 11 stages CMOS ring oscillator

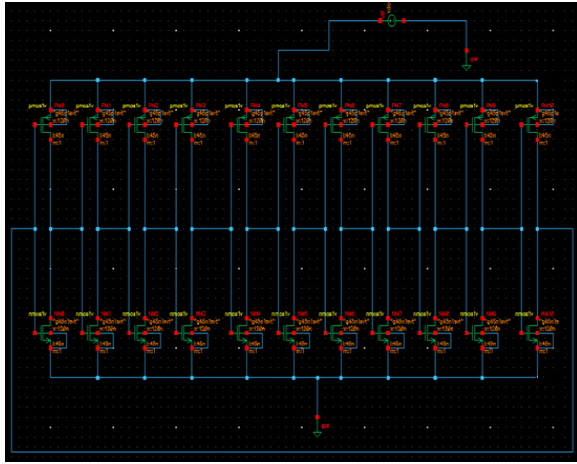


Fig. 2. Schematic diagram of 11 stages CMOS ring oscillator

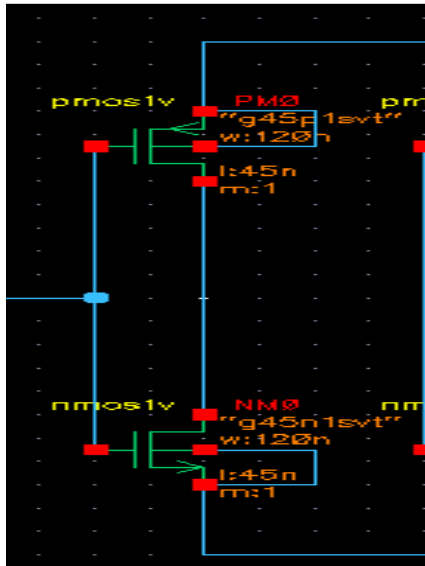


Fig. 3. One of the stages of 11 stages CMOS ring oscillator

Basically, for the performance analysis these circuits are simulated in spectre simulator of cadence tool using gpd045 library. The transistor sizing obtained as above for a set of

specifications are used to design the required ring oscillator circuit. The ring oscillator is designed for eleven stages of CMOS inverters and the W/L ratio which is 45nm CMOS Technology.

Table I. Design Parameters of 11 Stages Ring Oscillator

W_p (nm)	120
W_n (nm)	120
L (nm)	45

Actually, a circular chain composed of an even number of inverters cannot be used as a ring oscillator because the last output in this case is the same as the input. Although, this configuration of inverter feedback can be used as a storage element since it is the basic building block of static random access memory. A ring oscillator can be designed with a mix of inverting and non inverting stages provided the total number of inverting stages is odd [8]. Also, the oscillator period is equal to twice the sum of the individual delays of all stages. A time delay oscillator consists of an inverting amplifier with a delay element between the amplifier output and its input. The amplifier must have a gain of greater than 1 at the intended oscillation frequency [9]. A small amount of noise can cause the amplifier output to rise slightly. Having passed through the time delay element, this small output voltage change will be presented to the amplifier input [10]

A real ring oscillator only requires only power to operate above a certain threshold voltage oscillations begin spontaneously. By changing the supply voltage changes the delay through each inverter with higher voltages typically decreasing the delay and increasing the oscillator frequency. By Adding pairs of inverters to the ring increases the total delay and thereby decreases the oscillator frequency. [11].

3. RESULTS AND ANALYSIS

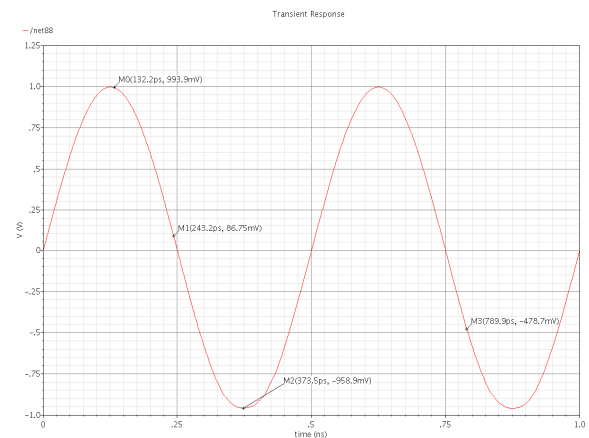


Fig. 4. waveform for Transient Analysis of 11 stages CMOS Ring Oscillator

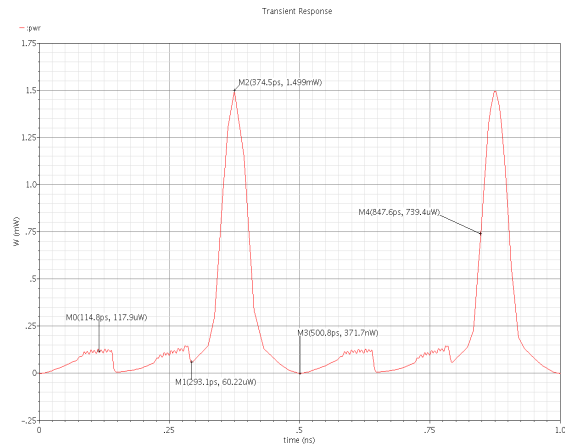


Fig. 5. waveform for Power consumption of 11 stages CMOS Ring Oscillator

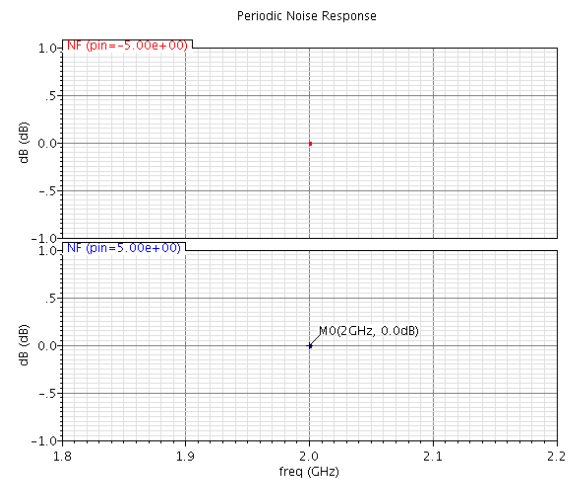


Fig. 9. waveform for Noise Figure of 11 stages CMOS Ring Oscillator

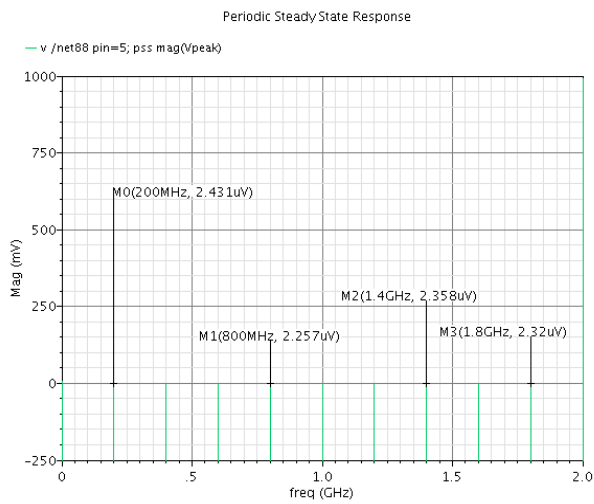


Fig. 6. waveform for Voltage of 11 stages CMOS Ring Oscillator

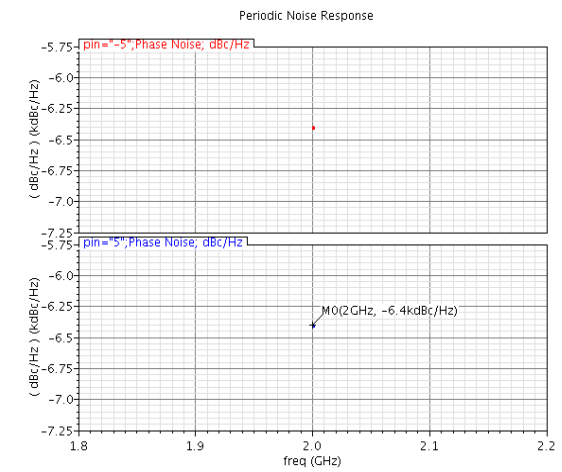


Fig. 10. waveform for Phase Noise of 11 stages CMOS Ring Oscillator

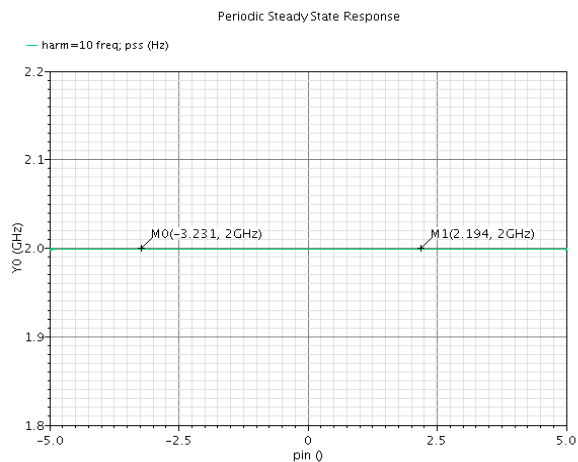


Fig. 8. waveform for Harmonic Frequency of 11 stages CMOS Ring Oscillator

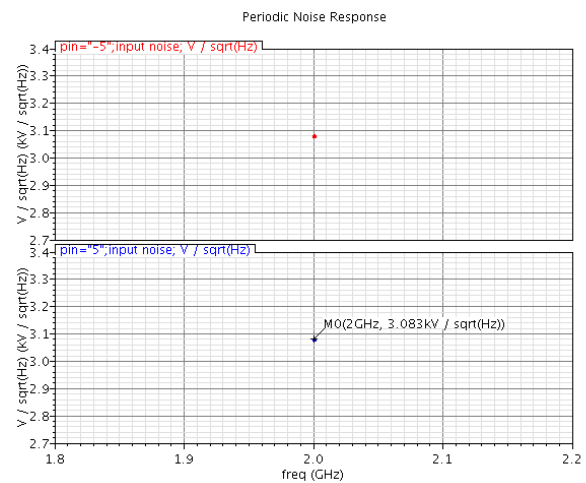


Fig. 11. waveform for Input Noise (10 dB) of 11 stages CMOS Ring Oscillator

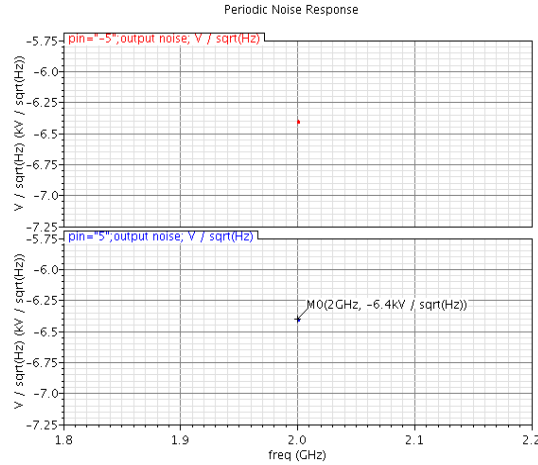


Fig. 12.waveform for Output Noise (dB 20) of 11 stages CMOS Ring Oscillator

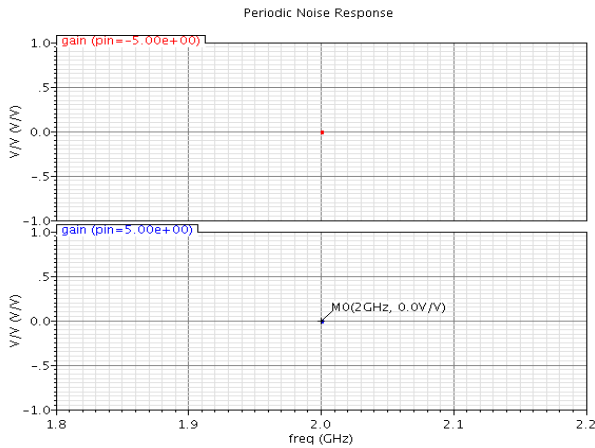


Fig. 13.waveform for Transfer Function of 11 stages CMOS Ring Oscillator

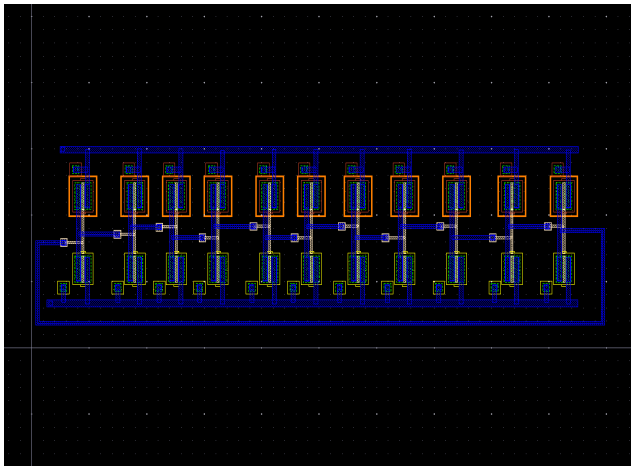


Fig. 14. layout of 11 stages CMOS Ring Oscillator in 45nm CMOS Technology

Table II.Performance Parameters of 11 Stages Ring Oscillator

Performance Parameters	Ring Oscillator with 11 stages
Average Power	5.702 mW
Frequency	2.0 GHz
Phase Margin	180°
Bandwidth	372.6 x10 ⁻¹²
Frequency Jitter	213.1 ps at 888.4 mV
Period Jitter	0.0 s at 749.7 ps
Voltage	2.257 μ V at 800 MHz
Harmonic Frequency	2 GHz at point 2.194
Phase Noise	-6.4 dBc/Hz at 2 GHz
Input Noise	3.083 kV/sqrt(Hz) at 2 GHz
Output Noise	-6.4 kV/sqrt(Hz) at 2 GHz
Noise Figure	0.0 dB at 2 GHz
Noise Factor	1 V ² /Hz at 2 GHz
Transfer Function	0 V/V at 2 GHz

A transient analysis (TA) has been performed to show the effects of the parameters over the oscillation frequency and to observe the behavior of the waveform with respect to time. A periodic steady state (PSS) analysis has also been performed to determine the frequency of oscillation and the influence of the parameters such as supply voltage, phase noise, timing jitter etc.

4. CONCLUSION

The eleven stages CMOS ring oscillator has successively been designed and executed using CMOS 45 nm technology process in Cadence virtuoso environment. Also, various performance parameters like noise, delay, glitches etc. were analysed and reduced accordingly to optimize the efficiency of the ring oscillator. Thus, the previous research work by the author has been improvised by obtaining the value of period jitter 0 at 933.5 ps, phase noise -6.4 dBc/Hz, Total Harmonic Distortion (THD) 3.586×10^{-4} at point -3.649.

REFERENCES

- [1] A. A. Abidi and R. G. Meyer, "Noise in relaxation oscillators," IEEE J. Solid State Circuits, vol. SC-18, no. 6, pp. 794–802, Dec. 1983.
- [2] B. Kim, D. Helman, and P. Gray, "A 30 MHz hybrid analog/digital clock recovery circuit in 2 μ m CMOS," IEEE J. Solid State Circuits, vol. 25, no. 6, pp. 1385–1394, June 1990.
- [3] J. Maneatis and M. Horowitz, "Precise delay generation using coupled oscillators," IEEE J. Solid State Circuits, vol. 28, no. 12, pp. 1273–1282, Dec. 1993.
- [4] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, Digital Integrated Circuits, Upper Saddle River, NJ: Pearson/Prentice Hall, 2003.
- [5] N. Weste and D. Harris, CMOS VLSI Design. Boston, MA: Pearson/Addison-Wesley, 2005.

-
- [6] R. Bayruns, R. Johnston, D. Fraser, and S. C. Fang, "Delay analysis of Si nMOSGbit/s logic circuits," *IEEE J. Solid State Circuits*, vol. 19, no. 5, pp. 755–764, May 1984.
- [7] N. Hedenstierna and K. Jeppson, "CMOS circuit speed and buffer optimization," *IEEE Trans. Comput. Aided Des. Integrated Circuits Systems*, vol. 6, no. 2, pp. 270–281, Mar. 1987.
- [8] L. Bisdounis, S. Nikolaidis, and O. Loufopavlou, "Propagation delay and short circuit power dissipation modeling of the CMOS inverter," *IEEE Trans. Circuits System International, Fundamental Theory Applications*, vol. 45, no. 3, pp. 259–270, Mar. 1998.
- [9] A. Kabbani, D. Al-Khalili, and A. Al-Khalili, "Technologyportable analytical model for DSM CMOS inverter transition time estimation," *IEEE Trans. Computer Aided Design Integrated Circuits Systems*, vol. 22, no. 9, pp. 1177–1187, Sep. 2003.
- [10] T. Weigandt, B. Kim, and P. Gray, "Analysis of timing jitter in CMOS ring oscillators," in *Proc. IEEE Int. Symposium Circuits and Systems (ISCAS)*, 1994, pp. 27–30.
- [11] A. Bell and G. Borriello, "A single chip nMOS Ethernet controller," in *IEEE International Solid-State Circuits Conference (ISSCC) Dig. Tech. Papers*, 1983, pp. 70–71.

Microcontroller Based Safety Guard System for Blind

Ruchi Gupta¹, Alka Verma², Okar Singh³, Renu Pooniya⁴

^{1,2}Department of EE & I Engineering
Moradabad Institute of Technology, Moradabad, 244001
¹ruchi.304@gmail.com, ²alkasinghmail@rediffmail.com

³Department of EC Engineering, Maharaja Agrasen College of Engg. & Tech., Amroha, 244231
omkar108@yahoo.com

⁴Department of EC Engineering, Shri Venkateshwara University, Amroha, 244231
renupunia622@gmail.com

Abstract: This paper demonstrates a prototype development of an Microcontroller Based Safety Guard System for blind people. Generally we see a lot of blind people around us who lead there live by depending on others. But with the help of MICROCONTROLLER BASED SAFETY GUARD SYSTEM FOR BLIND the conditions will definitely change. Now they will not have to depend on others for their survival instead they will be able to lead a self-dependent, respectful and an efficient life.

One of the basic and most fundamental concept with SAFETY GUARD SYSTEM FOR BLIND based upon microcontroller chips is that they are preprogrammed. Due to these preprogramming SAFETY GUARD SYSTEM FOR BLIND can sensed the signal the in real time from sensor and perform task according to the program which are already fixed. The proposed prototype systems is designed and demonstrated to recognize, understand and modify the actual performance and the movements of this done by getting information in real time from IR sensors connected to ATMEGA-89C51 microcontroller. A computer program is implemented in C-language and also controls this safety guard autonomously according to the received signals.

Keyword: ATMEGA-89C51, IR Sensor Microcontroller, Voice Recorder, Amplifier, Obstacle Detector.

1. INTRODUCTION

For the blind, it's difficult to step out without someone's help. To make the life simpler for them, here's an electronic safety guard system that alerts them of any obstacle or object in their path. The system can detect obstacles within one meter. The system comprises transmitter and receiver sections.

The transmitter is built around timer IC 555, which is designed to operate at a frequency of 38 KHZ. This signal is amplified by a current amplifier and transmitted through infrared (IR) diodes. The receiver section consists of an IR sensor TSOP1738, power amplifier, comparator IC LM311,

microcontroller IC AT89C51, relay driver and voice processor IC ISD1420. The IR rays reflected back from any obstacle are received by the IR receiver. The received signal is amplified by the amplifier stage, so even the weak signals can be picked up by the receiver. The amplified signal voltage is compared with a fixed threshold voltage at comparator LM311. The output of the comparator is given to the input/output (I/O) port of the microcontroller. When the comparator output goes high because of reflection of signal from an object, the microcontroller energizes a relay via the relay driver. The relay contacts are used by a voice processor to play a prerecorded warning message (such as "hey, there's an obstacle"). The user can hear the played message using a headphone.

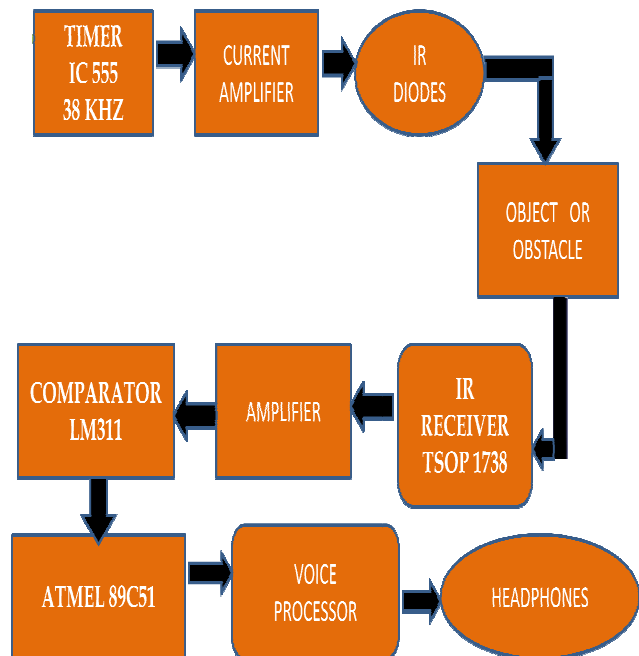


Fig.1. Basic Building Block Of Obstacle and Edge Detector Robot

A. Project Objectives

- The main goal of the project is to provide a cost-effective way to allow buildings to support blind people.
- The Blind Audio Guidance System hopes to allow visually impaired users to simply press a button, speak the desired destination, and be guided there with the use of audio instructions.
- The system hopes to provide a portable unit that can easily be carried and operated by a visually impaired user. It could easily be incorporated into a walking cane

2. CIRCUIT DESCRIPTION

A. Transmitter Section

The transmitter circuit is powered by a 9V battery. When switch S1 is closed, LED1 glows to indicate the presence of power in the circuit. Timer IC555 (IC1) is wired as an astable multivibrator. The output frequency (38 KHZ) of IC1 at its pin 3 can be varied using VR1 (2k). The output of IC1 is given to the base of npn transistor T1 and T2 (each BC548) form a Darlington pair that boosts the output current to drive the two infrared diodes connected in series at the collector of the Darlington pair (IR LED1 and IR LED2). The output signal frequency of 38KHZ is transmitted by the IR LEDs.

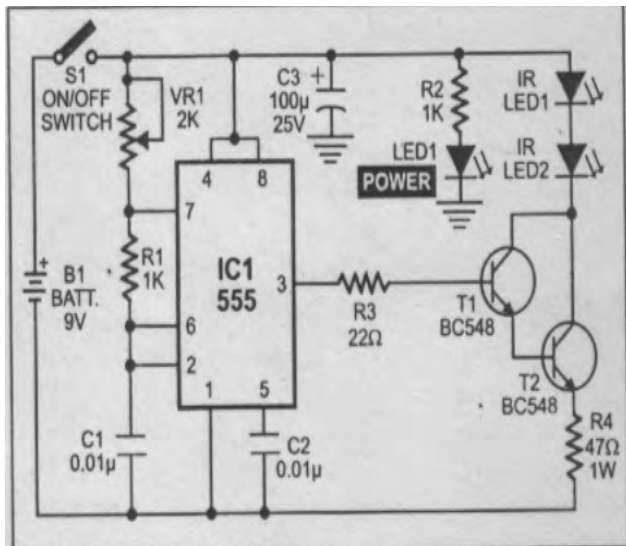


Fig. 2. Transmitter Circuit Diagram

B. Receiver Section

The receiver circuit is powered by a 9V battery. When switch S2 is closed, LED2 glows to indicate the presence of power in the circuit. The 9V supply is down-converted to 5V using regulator IC 7805 (IC2) to drive the IR receiver

module (TSOP1738), microcontroller and voice processor sections. The IR rays path of the user is received by the IR receiver module. This signal is amplified by the power amplifier stage comprising transistors T3, T4 and T5 (each BC548). The amplified output at the emitter of transistor T5 is given to the non-inverting input (pin 2) of comparator IC LM311 (IC3) through resistor R13.

A reference voltage of 2.2V developed across Zener diode ZD1 is connected to the inverting input (pin 3) of IC3. When the voltage level at pin 2 increases beyond the reference voltage, output pin 7 of IC3 goes high, this is indicated by the glowing of LED3. This output is given to the I/O port P1.0 of microcontroller IC4

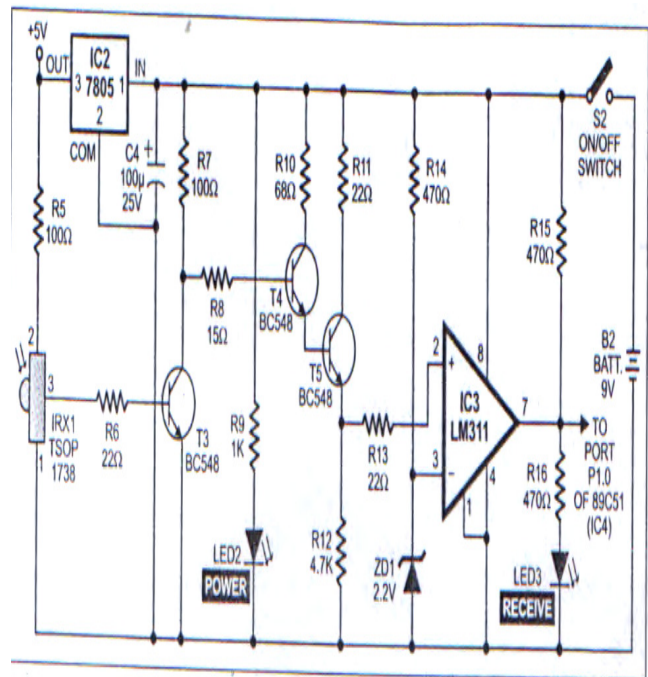


Fig. 3. Receiver Circuit Diagram

C. Microcontroller Section

Microcontroller chip AT89C51 (IC4) acts as a switching hub only and can be replaced by any other switching circuit. The use of this chip in this circuit is to show how to interface an embedded system in a home made project the program burnt into this chip decides the action when a signal is received at its input. As shown in Fig., ports P1.0 through P1.3 of IC4 are used as the input ports. The corresponding outputs are available at ports P2.0 through P2.3. The output of the comparator is fed to port P1.0 and the corresponding output at port P2.0 is fed to the base of transistor T6 (BC558) through resistor R18. Normally, when no signal is applied at input port P1.0, output P2.0 is high. When input P1.0 becomes high, output P2.0 goes low and transistor T6 conducts. This, in turn, drives transistor T7 (BC548) to

energises relay RL1, which is indicated by glowing of LED4. In case the circuit behaves abnormally, presses switch S3 momentarily to rest the circuit.

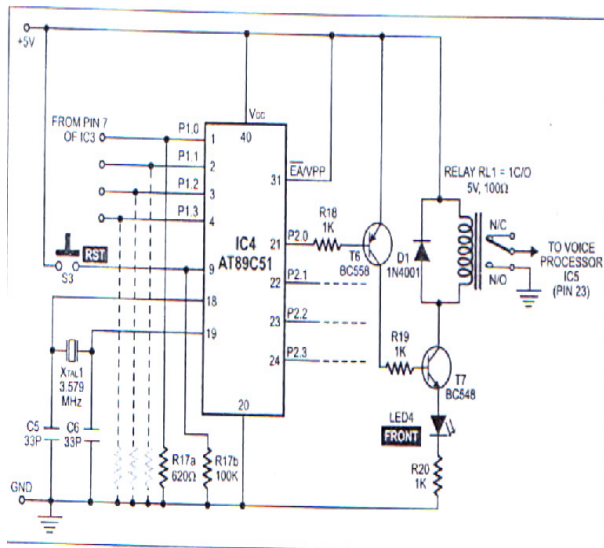
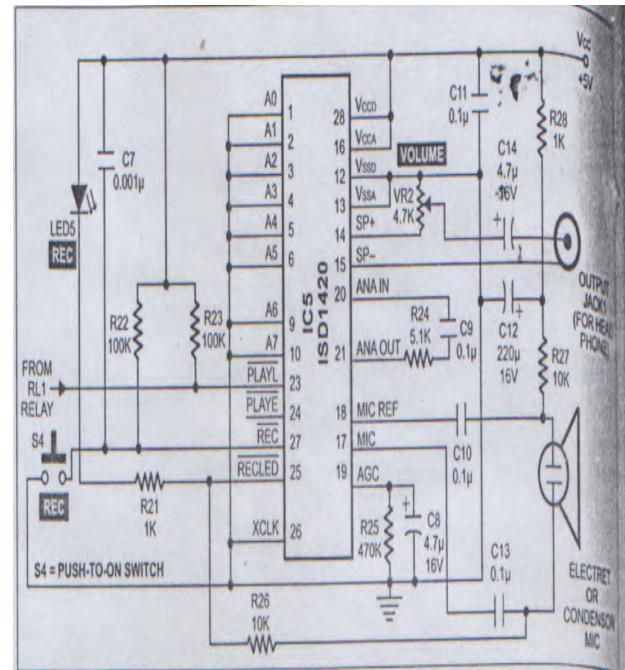


Fig. 4. Microcontroller Circuit

D. Voice recorder Section

The voice processor section receives regulated 5V DC supply section receives regulated 5VDC supply from regulator IC2. Voice processor IC ISD1420 (IC5) used here is a 28-pin chip from Win bond. It can record a voice message can be played at the press of a button connected to one of its pins. The pushbutton switch S4 connected to pin 27 of IC5 is used for recording the message in the processor. Pin 23 is used for playing the recorded message. The condenser microphone for inputting the voice message is connected to pins 17 and 18 of IC5 via capacitors C13 and C10, respectively. The message is output via pins 14 and 15. A loudspeaker or headphone can be directly connected to these pins through a coupling capacitor. Here, we've used an output jack (JACK1) at these pins for headphone connection. Preset VR2 is used to control the volume and C14 acts as a coupling capacitor. Keep switch S4 pressed (maximum for 20 seconds) as you speak into the microphone for recording the message. Releases switch S4 after recording is done. To listen to the recorded message through the speaker or the headphone, playback pin 23 (PLAYL) must be held down to ground. Here, energisation of relay RL1 pulls pin 23 to ground and thus enables playback of the recorded message. The pole of relay RL1 is connected to pin 23 of IC5, while the normally open (N/O) contact is grounded. When relay RL1 energises, the pole of the relay connects to the N/O contact enabling the voice processor to play the recorded message and the message can be heard from the headphone.



4. FUTURE SCOPE

IR sensors can be used to automatically detect and avoid obstacles if the robot goes beyond the line of sight. This avoids damage to the vehicle if we are operating it from a distant place. If the current project is interfaced with a camera, robot can be driven beyond line of sight and range becomes partially unlimited as GSM network has a very large range.

5. CONCLUSION

The project “MICROCONTROLLER BASED SAFETY GUARD FOR BLIND” had been proven to be challenging and it had given us the opportunity to experience the entire process of building a needful microcontroller project. This project aims on making the life of blinds simpler and easier. To sum it up further, this project are divided into the hardware and software part. Both the parts are equally important as it need each other to operate and to reach its objectives. The hardware part can be mainly divided into three main parts, namely, the Transmitter system, the microcontroller and the Receiver system. The Transmitter system has the job to transmit infrared signal, the receiver has the job to receive the infrared signal and send it to the microcontroller. Using the data from the receiver, the microcontroller activates the voice processor through reset switches. Once the receiver received the data, its job is to send the signal to the microcontroller to make it audible to the user. Even though this project has certain limitations, it was proven that it can achieve its objective by sending the signal to the microcontroller. The transmitter, receiver and the microcontroller can be mounted on a stick for blind and may be used easily. The software parts are mainly of C programming software, which is a product of Keil software developers. This software helps the user to program the microcontroller easily and within few lines. The Interface for this project was created using the C language. There are some shortcomings in this project. One of the obvious one is the fact that the stick cannot sense faraway objects. It was actually planned to use the higher sensitivity components in the beginning but due to shortage and limitation in budget, it could not be done. All in all, the project is a success in term of the objective, as it manages to reach the objective that

was set. Though the result is satisfactory, it would have been a much better result if the range of the transmitter and receiver was more, as it would help the user to sense objects from a larger distance. This project needs a lot of improvement, and further suggestions will be surely welcomed.

6. ACKNOWLEDGEMENT

We express our deepest sense of gratitude to Mrs. ALKA VERMA. She deserves special thanks for her constant encouragement, invaluable guidance and faith in us.

REFERENCES

- [1] Howard, M.-J. Marari'c, and G.-S. Sukhatme (2002), An incremental self-deployment algorithm for mobile sensor networks, *Autonomous Robots*, Vol. 13, pp. 113-126.
- [2] L. Li, J.-Y. Halpern, P. Bahl, Y.-M. Wang, and R. Wattenhofer (2001), Analysis of a conebased distributed topology control for wireless multi-hop networks, in *Proceedings of the Twentieth Annual ACM Symposium on Principles of Distributed Computing*.
- [3] F. Mondinelli and Z. K. Vajna (2002), Self localizing sensor network architectures, in *Proceeding of the 19th IEEE Instrumentation and Measurement Technology Conference*.
- [4] E. L. Lloyd, R. Liu, M. V. Marathe, R. Ramanathan, and S. S. Ravi (2002), Algorithmic aspects of topology control problems for ad hoc networks, in *Proceedings of the 3rd ACM International Symposium on Mobile Ad-Hoc Networking and Computing*.
- [5] H. Gupta, S.-R. Das, and Q. Gu (2003), Connected sensor cover: Self-organization of sensor networks for efficient query execution, in *Proceedings of the 4th ACM International Symposium on Mobile Ad-Hoc Networking and Computing*.
- [6] G. Gupta and M. Younis (2003), Load-balanced clustering of wireless sensor networks, in *Proceedings of the IEEE International Conference on Communications*.
- [7] W.-R. Heinzelman, A. Chandrakasan, and H. Balakrishnan (2000), Energy-efficient communication protocol for wireless micro sensor networks, in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*.
- [8] C.-Y. Chang, C.-C. Hsieh, and H.-R. Chang (2005), Obstacle-free robot deployment algorithm for wireless sensor networks, in *Proceedings of the Workshop on Wireless, Ad Hoc, and Sensor Networks (WASN 2005)*.

Heuristic Algorithm Approach for FSM Decomposition to Reduce Power

Himani Mittal¹, Naved², Predeep Kumar³

¹Assistant Professor in J.S.S.Academy of Technical Education, NOIDA
himanimit@yahoo.co.in

²Associate Professor in G.B.U, GautamBudh Nagar

³Professor in Amity University, NOIDA

Abstract: Power consumption and power-related issues have become a major concern for most designs. The primary method used for reducing power is supply voltage reduction, this technique begins to lose its effectiveness as voltages drop to sub threshold volt range and further reductions in the supply voltage begin to create more problems than are solved. We present in this article a new approach to the synthesis problem for finite state machines with the reduction of power dissipation as a design objective. A finite state machine is decomposed into a number of coupled sub machines. Most of the time, only one of the submachine will be activated which, consequently, could lead to substantial savings in reduction power consumption. A low power architecture for designs modeled as an FSM. It is based on the concept in which the redundant computation can be dynamically disabled to reduce the overall power dissipation.

Keywords: FSM Decomposition[2], Mealy and Moore Machines, STM (State Transition Matrix), ULP (Ultra Low Power), Power saving

1. INTRODUCTION

In CMOS circuits, power is dissipated in a gate when the gate output changes from 0 to 1 or from 1 to 0. Minimization of power dissipation can be considered at algorithmic, architectural, logic, and circuit levels. In sequential circuit design, an effective approach to reduce power dissipation is to “turn off” portions of the circuit, and hence reduce the switching activities in the circuit. In this article we propose a technique that is also based on selectively turning off portions of a circuit. Our approach is motivated by the observation that, for an FSM, active transitions occur only within a subset of states in a period of time. Therefore, if we synthesize an FSM in such a way that only the part of the circuit which computes the state transitions and outputs will be turned on while all other parts will be turned off, power consumption will be reduced. In a CMOS circuit, generally, the switching activity of the gate output contributes most to the total power dissipation. For FSM low power design, partitioning technique proves to be effective for reducing switching activity. That is, partition the original FSM into several smaller sub FSMs and only one of them is active at a time. However, two issues are often introduced:

- (1). Increased area overhead of state memory by separate sub FSM encoding.
- (2). More power consumption by crossing transitions between different sub FSMs in the synchronous Design

Objective of this paper is :

- (1) Computing the Power consumption in original FSM & after decomposition approach computing power consumption in sub FSMs.
- (2) Comparing the consumption of power in original FSM and decomposed FSMs.
- (3) To develop an architecture of Low Power FSM & computing power consumption at architecture level.

2. METHODS AND APPROACHES

The key steps in our approach are:

- (1) Decomposition of a finite state machine into submachine so that there is a high probability that state transitions will be confined to the smaller of the sub machines most of the time.
- (2) Synthesis of the coupled sub machines to optimize the logic circuits i.e. to assign state codes to the states of the sub machines.
- (3) It is based on the concept in which the redundant computation can be dynamically disabled to reduce the overall power dissipation.
- (4) When the input signal arrives, the active submachine remains active. In that case, the other

Submachine will not be turned on and remain inactive. On the other hand when next input signal arrives, the active submachine might turn on another submachine and turn itself off, becoming inactive.

- (5) At any moment, only one submachine is active (with its corresponding combinational circuit turned on) while all other sub machines are inactive (with their corresponding combinational circuits turned off).
- (6) An effective approach to reduce power dissipation is to “turn off” portions of the circuit, and hence reduces the switching activities in the circuit. We synthesize an FSM in such a way that only the part of the circuit which computes the state transitions and outputs will be turned on while all other parts will be turned off.
- (7) In general, since the combinational circuit for each submachine is smaller than that for the original machine, power consumption in the decomposed machine will be smaller than that of the original machine.

3. BASIC PRINCIPLES

Entropy is a measure of the randomness carried by a set of discrete events observed over time. In the studies of the information theory, a method to quantify the information content C_i of an event E_i in this manner is to take logarithmic of the event probability

$$C_i = \log_2 (1/P_i)$$

Since $0 \leq P_i \leq 1$, the logarithmic term is non negative and we have $C_i > 0$.

The average information contents of the system is the weighted sum of the information content of C_i by its occurrence probability. This is also called the entropy of the system.

$$H(X) = \sum_{i=1}^{m-1} p_i \log_2 \frac{1}{p_i}$$

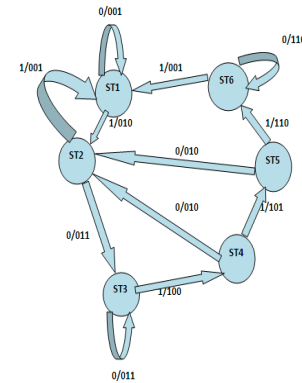
4. SHORTCOMING OF THE SOLUTION

The major shortcoming of this solution, however, is that lowering the supply voltage affects circuit speed. As a consequence, both design and technological solutions must be applied in order to compensate the decrease in circuit performance introduced by reduced voltage. In other words, speed optimization is applied first, followed by supply voltage scaling, which brings the design back to its original timing, but with a lower power requirement. A similar problem, i.e., performance decrease, is encountered when power optimization is obtained through frequency scaling. Techniques that rely on reductions of the clock frequency to lower power consumption are thus usable under the constraint that some performance slack does exist. Although this may seldom occur for designs considered in their entirety, it happens quite often that some specific units in a larger architecture do not require peak performance for some

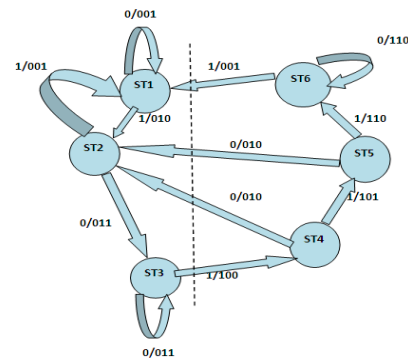
clock/machine cycles. Selective frequency scaling (as well as voltage scaling) on such units may thus be applied, at no penalty in the overall system speed. Optimization approaches that have a lower impact on performance, yet allowing significant power savings, are those targeting the minimization of the *switched capacitance* (i.e., the product of the capacitive load with the switching activity). Static solutions (i.e., applicable at design time) handle switched capacitance minimization through area optimization (that corresponds to a decrease in the capacitive load) and switching activity reduction via exploitation of different kinds of signal correlations (temporal, spatial, spatial-temporal). Dynamic techniques, on the other hand, aim at eliminating power wastes that may be originated by the application of certain system workloads (i.e., the data being processed).

5. DECOMPOSITION APPROACH

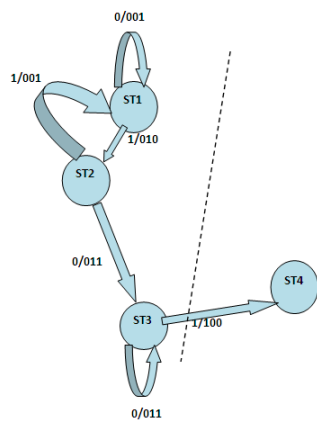
1. Keep minimum crossing Transition to reduce power consumption.[2]
2. First bit of state code is Control bit distinguish between Sub machines.
3. Inner two bits are to distinguish between states within each submachine.



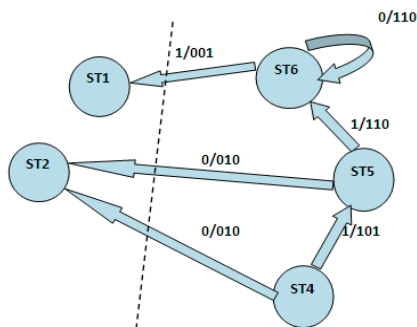
Original FSM



Partitioning of FSM



Upper half FSM



Lower half FSM

6. SIMULATION

After decomposing this original FSM and simulated it using MODELSIM[1] following waveforms are observed .

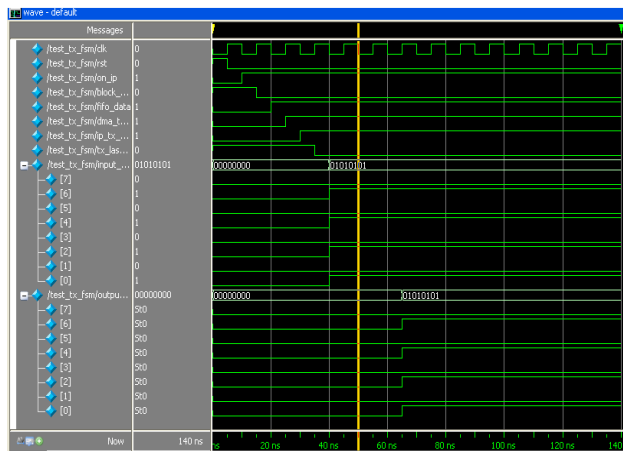


Fig. (a) Output Waveform representing State Transition in FSM for an arbitrary input sequence.

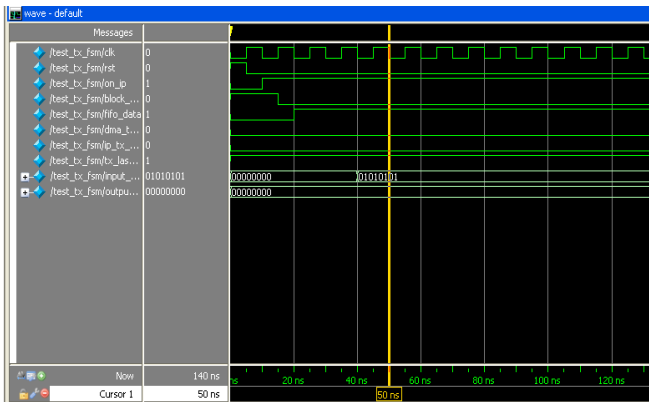


Fig. (b) Output Waveform representing State transition in Upper half FSM.

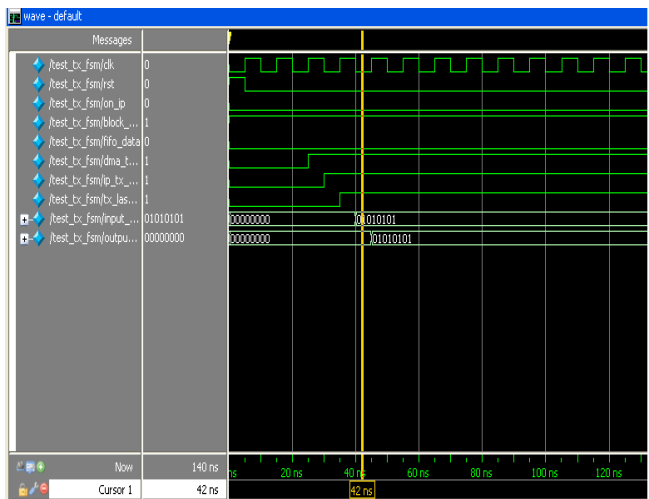


Fig. (c) Output Waveform representing State transition in Lower half subFSM

7. CALCULATION AND RESULT

Power consumption in original FSM is calculated as 6.08mW.

Power consumption in decomposed sub FSM is 4.7mW & 3.3mW.

On taking the average, 4.064mW.

Percentage in power saving[3] = $(P_{\text{originalFSM}} - P_{\text{subFSM}})$

$$P_{\text{originalFSM}}$$

$$= 33.13\%$$

8. CONCLUSION

The Decomposed FSM[2] technique leads in a 33.13% average reduction in switching activity of the state variables, and 12% average reduction of the total switching activity of

the implemented circuit. Although the solution is heuristic, and does not guarantee the minimum power consumption, these results leads to a reduction in the power consumption in the complete circuit. Improved methods of decomposing FSM in more than two subFSM based on control bits considering more number of states and decomposing it into three subFSMs and comparing the saving in Power consumption with original FSM. Use the tool for "real-world" designs in the areas of microprocessors, data communication, and signal processing.

REFERENCES

- [1] J. Liang, Swaminathan, "A SOC: a Scalable, Singlechip Communications Architecture", Tessier, R. Parallel Architectures and Compilation Techniques, 2000, Proceedings. International Conference, pp.37-46.
- [2] Guoliang Ma, Hu He, "Design and Implementation of an Advanced DMA Controller on AMBA-Based SoC", IEEE 8th International conference, 2009, pp 419-422
- [3] Flynn, D., "AMBA: Enabling Reuseable On-Chip Designs", IEEE Micro, 17(4), July/August 1997, pp 20-27.
- [4] AMBA Specification (rev2.0) and Multi Layer AHB Specification, Arm: <http://www.arm.com>, 2001.
- [5] Primecell DMA Controller (PLO80) Technical Reference Manual, <http://www.arm.com>
- [6] C8237 Programmable DMA Controller Core, <http://www.cast-inc.com>
- [7] A. Bogliolo, and G. De Micheli, "A survey of Design Techniques for System-Level Dynamic Power Management," IEEE Trans . On VLSI Systems, Vol.8, No.3, pp, 299-316, June 2000.
- [8] A. Chandrakasan and R. Brodersen, Low- Power CMOS Design. IEEE Press, 1998.
- [9] A. Macii, E. Macii, L. Benini, and M. Poncino, "Selective Instruction Compression for Memory Energy Reduction in Embedded Systems", ISLPED-99: ACM/IEEE 1999 International Symposium on Low Power Electronics and Design, pp. 206-211, San Diego, California, August 1999.
- [10] A. Despain, C-Y. Tsui, J. Monteiro, M. Pedram, S. Devadas, and B. Lin. Power Estimation Methods for Sequential Logic Circuits. IEEE Transactions on VLSI Systems, 3(3):404-416, September 1995

Independent Control of Rise and Fall Edge Dead Time Using IC 555 Timer

Ritish Kumar¹, Jitender Kumar²

¹Meerut Institute of Engineering & Technology, Meerut, India

²Ambedkar Institute of Advanced, Communication Technologies and Research, Delhi, India

¹kr.ritish@gmail.com, ²jeet.kumar0407@gmail.com

Abstract: In this paper we study the circuit in which rise and fall edge dead time can be controlled independently and this can achieve upto nanosecond of dead time level. The condition for the circuit to work properly is, the two dead times (t_1 , t_2) must be smaller than half of the time period of applied input pulse. If the circuit is following this condition then variation in dead time will only affect duration of ON and OFF states of the output signals, there will be no effect of dead time variation on total time period of output signals and total time period of two output signals will remain same as time period of input pulse. Because of nanosecond dead time level this circuit can operate at high frequency. Furthermore, circuit is easily implemented. It hence can be employed in various applications.

Keywords: rise and fall edge; dead time; 555 timer

1. INTRODUCTION

The scheme of circuit is based on monostable multivibrator circuit using IC 555 timer and the toggle circuit that uses JK flip-flop to operate. The dead time at rise and fall edge can be controlled independently and this can be achieved by adjusting the values of resistor and capacitor that are combined in multivibrator circuit. This circuit is easily implemented and can provide precise dead time duration in nanosecond. Because of nanosecond dead time, it can operate at high frequency effectively and can be employed in high frequency applications.

Many electronic systems involve with two electronically controllable switches. In most cases they are controlled in a complementary manner. In case that both switches simultaneously on even for short period, this lead to short circuit, heating and destruction of circuit components. Dead time duration that is inserted between two switching control signals is the precaution against this effect. It is found that dead time duration may cause distortion in an output of inverter/ converter, instability of the system including distortion in class-D amplifier or pulse dropping in PWM. As the result, dead time should be properly defined which depends on circuit parameters, operation conditions and applications.

Many techniques were proposed to achieve optimum dead time duration. In most cases inserted dead time is usually of

a fixed duration for easy implementation. But fixed dead time is not suitable for some applications, hence adjustable dead time is required. There are techniques for adjustable dead time but the drawback is, they can only provide dead time duration of microsecond. Slow recovery of monostable multivibrator is also a big disadvantage which can cause dead time fluctuations at high counting rate. In many other techniques designs are only focused on particular application and provide a single symmetric rise and fall edge dead time and because of microsecond dead time they cannot efficiently operate at high frequency conditions. This paper thus presents the independent rise and fall edge dead time generator with a circuit which provides very fast recovery characteristics. In addition, rise and fall edge dead time duration can be independently adjusted in nanosecond level and the circuit can operate at high frequency. Furthermore, the proposed circuit can be implemented easily and can be effectively applied in various applications.

2. PRINCIPLE

The circuit for generating variable dead time duration is based on monostable multivibrator using 555 timer with JK flip-flop in toggle form and logic gates as shown in Fig.1.

From Fig.1., the input signal be square wave as shown in Fig.2(S1), which is triggered at rise edge and fall edge by monostable multivibrator circuits and two variable pulse width (t_1 and t_2) outputs are obtained as shown in Fig.2(S2) and Fig.2(S3) respectively. The value of t_1 can be equal to or it can be different from t_2 , it depends on the values of resistors and capacitors of multivibrator circuits. As we can see from Fig.2, for the circuit to work properly time period of both multivibrator generated pulses (t_1 , t_2) must be smaller than $T/2$ (half of the total time period of the input square pulse). Both the pulses are ORed and resulted signal is shown in Fig.2(S4). The signal from OR is toggled at rise and fall edge by toggle circuit which is constructed using JK flip-flop. The output signals are shown by Fig.2(S5) and Fig.2(S6) respectively. By feeding both output signals from toggle circuit into OR gate, it yields the signal as shown in Fig.2(S7) and Fig.2(S8) after be inverted by NOT gate. Similarly, another output signal shown in Fig.2(S9) is

obtained by feeding both output signals from toggle circuit into AND gate.

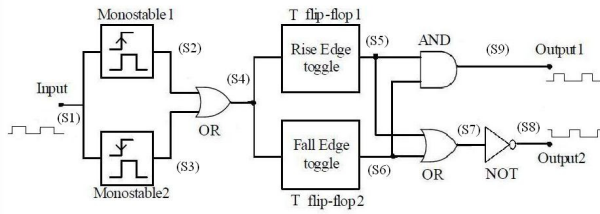


Figure 1. Block Diagram of proposed circuit.

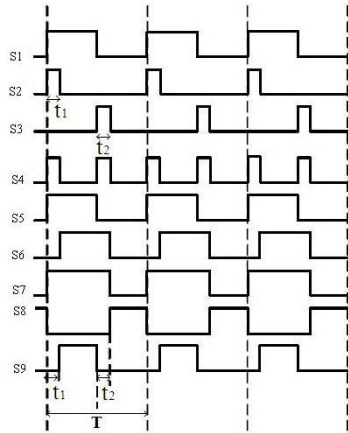


Figure 2. State diagram of input and output voltage in time domain.

From Fig.2, it is found that both output signals will not be in high state simultaneously. The dead time period can be defined as

$$T_{\text{dead time}} = t_1 + t_2$$

Where t_1 and t_2 are the output pulse widths of monostable multivibrator circuits which triggering at rise and fall edge of the input signal respectively. These pulse widths depend on resistor R_1 , R_2 and capacitor C_1 , C_2 that are combined in each monostable multivibrator circuit. Pulse width can be calculated by

$$t_1 = 1.1 R_1 C_1$$

$$t_2 = 1.1 R_2 C_2$$

3. EFFECTS OF DEAD TIME

As we can see from Fig.2(S9) and Fig.2(S8) the two outputs of the circuit are not in high state simultaneously at any instant of time. So, we can use this circuit where we need to

control a pair of switches in alternate manner. To avoid a short circuit situation (both signals in high state simultaneously) we have inserted dead time t_1 (between fall edge of output2 and rise edge of output1) and t_2 (between fall edge of output1 and rise edge of output2). Using these dead times we can control ON and OFF duration of the two operating switches. Also we can see from Fig.2 that inserting dead time t_1 we are getting output1 for which ON time will be t_1 less than ON time of input pulse. Similarly, we can say inserting dead time t_2 will give us output2 which is inverted version of the pulse for which ON time is t_2 more than the ON time of input pulse. So, we can say that any variation in dead time will only affect ON and OFF states of the output signals and total time period of output signals will remain same that is equal to the time period of input pulse.

4. CONCLUSION

Here we have generated rise and fall edge dead time which can be controlled independently without affecting the dead time of other edge. Since the circuit is providing nanosecond dead time level with an ability to operate at high frequencies and it also has very fast recovery option so we can use it in many applications such as PWM controlled inverter/converter or system, modulation strategies for class-D amplifiers, MOSFET driving for synchronous converter and rectifiers. The precaution to be use here is, we have to take care that pulse widths of multivibrators' outputs (t_1 , t_2) must be smaller than half of the time period of the input pulse. If the circuit is working properly then any variation in t_1 and t_2 will only affect period of ON and OFF states and total time period of output signals will remain same as time period of input pulse.

REFERENCES

- [1] B. Zhou, W.H. Lau and H. Chung, "The Analysis of a Novel Dead-Time Generation and Compensation method for 2-level PWM Topology," IEEE Power Electronics, PESC'06, 37th, 2006.
- [2] L. Yong-Kai, and L. Yen-Shin, "Dead-Time Elimination of PWM Controlled Inverter/Converter without Separate Power Sources for Current Polarity Detection Circuit," IEEE Trans. On Industrial Electronics, Vol. 56, No.6, June 2009.
- [3] J. Seung-Gi, and P. Min-Ho, "The analysis and compensation of dead time Effect in PWM inverters," IEEE Trans. On Industrial Electronics, Vol. 38, No. 2, April 1991.
- [4] J.M. Schellekens, R.A.M. Bierbooms, and J.L. Duarte, "Dead-Time Compensation for PWM Amplifiers using Simple Feed-forward Techniques," XIX International Conference on Electrical Machines - ICEM 2010.

Spatial Filters: A FSS Approach of Designing

Vishakha Kanwar¹, Ankush Kapoor²

¹B.Tech 2nd Year ECE Department

Jawaharlal Nehru Government Engineering College, Sundernagar Distt. Mandi H.P, Pin-175018

vishakha.kanwar6@gmail.com

²Assistant Professor ECE Department, Jawaharlal Nehru Government Engineering College

Sundernagar Distt. Mandi H.P, Pin-175018, ankush8818@yahoo.com

Abstract: This paper presents the importance and necessity of frequency selective surface filters. The main purpose of this paper is to give an idea about spatial filter for communication and radar system. We have simulated the square loop FSS structure and we have seen the effects of varying the substrate thickness on its performance. It consists of metallic grid deposited on a polymer substrate. It is a two dimensional periodic structure which has the characteristics of bandpass when interacting with electromagnetic waves. When an electromagnetic wave is incident on the surface at a wide angle, an optimised frequency selective surface structure and its transmission characteristics are obtained. They are also known as filters of electromagnetic waves, mainly because of their function of band blocking. Currently FSS are in use as 'RADOMES' for antennas and controlling their radar cross-section.

Keywords: Frequency Selective Surface (FSS), Spatial Filters, Metamaterials, Radomes.

1. INTRODUCTION

Frequency selective surface (FSS) is a type of a spatial filter. Basically a filter can be used in any electronic device or in any other application which will control the content of frequency of the signal for discriminating noise and unwanted interference [1-2]. When electromagnetic radiation is incident on frequency selective surface then it act as a spatial filter, where some frequency band will then be transmitted and some will be reflected. Research on FSS began between the late 60s and the early 70s, while the periodic electromagnetic bandgap structure has been investigated increasingly in this era. Research on FSS began in China on the 1990s, concentrating mainly on the application of band pass to radome. Now the study is involved with the application of the FSS to wave absorption and antenna technology.

Frequency behaviour of FSS is entirely determined by the geometry of the surface in one unit cell when surface size is extending upto infinity [2]. The various parameters to be considered while designing frequency selective surface are the shape, size, angle dependence, bandwidth and band separation. A frequency selective surface is a periodic planar assembly made up of made up of metallic elements on a

dielectric layer [3]. Periodic structure arrangement or we can say the structure having equal space between them is having an important role in physics and engineering. The periodic structure is used in the designing of FSS which encounters certain problems:-

- (1) In which pattern of the periodic structure is to be determined.
- (2) The problems in which the parameters of a fixed pattern are to be determined.

In order to solve above problems, the element type is assumed to be square loop, and the approach is to adjust the parameters such as element size and spacing for desired feature. Generally FSS is placed on a dielectric or insulated material which is to be carefully chosen. Also the material used to manufacture the frequency selective surface is chosen on the basis of different parameters based on the criteria that it shouldn't affect performance for which it is to be meant [4]. According to the FSS theory, more than one solution might be found depending on the type of FSS (Inductive or Capacitive), type of single-cell element (crossed-dipoles, dual-rings, etc) and FSS metal depth. Different structures of the conducting elements (or apertures) lead to various resonant characteristics. Also the choice of the substrate is the main area of concern to get the desired response. The effects of substrate parameters on the FSS are discussed in the literature [5, 6, 8].

FSS act as a microwave filter which is a two-port network used to control the frequency response at certain point in a microwave system by providing transmission at frequencies within the passband of the filter and attenuation in the stop band of the filter[12-16]. We used the Thermocol material and studied the transmissive and reflective properties by varying the thickness of the material on the CST software.

The organization of the manuscript is as follow. Section 2 deals with the Metamaterials and in Section 3, the Experimental results obtained by using the CST Simulation software were presented. Finally, Section 4 concludes the work.

2. METAMATERIAL: A NEW APPROACH

Electromagnetic bandgap metamaterials can be used to control the amount of light propagating. This can be done by 'Photonic crystals' or Left handed materials, which are the types of metamaterials. Metamaterials can be used as a element in FSS, because it can enhance the radiated power of antenna. As metamaterials is having negative permeability and permittivity therefore it shows the properties like small size of antenna, improved efficiency and bandwidth performance. By designing FSS with metamaterial allows smaller antenna elements to cover a wider frequency range. Various metamaterial systems can be used to support the surveillance sensor, center of communication, navigation systems, control or commanding systems [10]. The properties of metamaterials are not found in nature. These are made artificially. These type of antennas are getting preference because these have the properties of discriminating the limitations of bandwidth for the natural or conventially constructed electrically small antennas.

Metamaterials, having high impedance ground planes can also be used to improve the radiation efficiency and it also improves the radio performance of low profile antennas. These can also be used to enhance the scanning range of the beam with the help of forward and backward waves in leaky wave antennas. We can conclude from the above discussion that metamaterial have entered in the field of electromagnetic and are really advantageous for the designing of FSS.

3. DESIGN APPROACH

In this work we have analyzed several techniques which were previously being used to develop the FSS and chosen Equivalent Circuit Model approach.

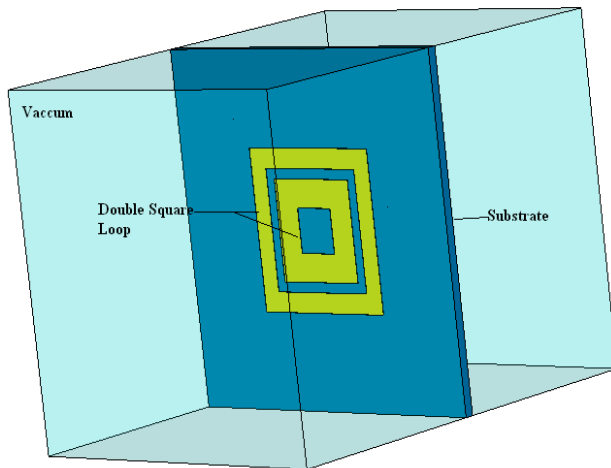


Fig. 1. Double Square Loop structure.

In this approach we have made use of the two robust and widely chosen software packages which are CST and An soft Designer. Using circuit simulation software. An soft Designer we can get the response and study the transmissive and the reflective properties. CST is really a asset in the real time simulation of the Electromagnetic structures and provide accurate real time transmissive and the reflective properties.

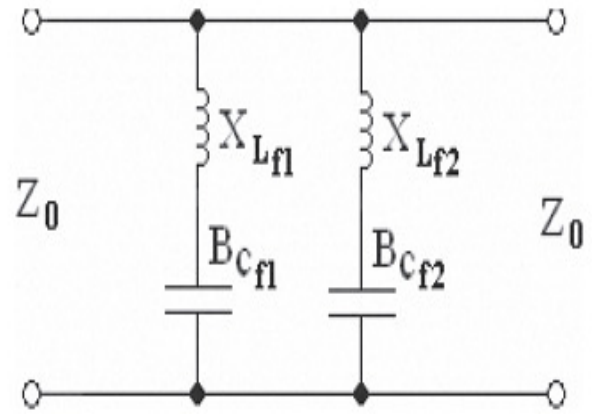


Fig. 2: Equivalent circuit model [4].

In the above Fig. 1 and Fig. 2 the structure and its equivalent circuit which we have used is clearly visualized. Equivalent Circuit Model offers a simple mathematical approach in FSS analysis and it also requires limited computational resources when compared with other methods, so it helps to quickly predict the performance of the structures. On the basis of transmission line analogy the interaction of incident waves with an FSS is represented as a wave travelling down a transmission line, with shunt lumped circuit impedances.

The modelling technique for the development of equivalent circuits for FSS structures is based on equations given by Marcuvitz [11], who developed the initial expression for the periodic gratings. Generally, it must be true also as the substrate is none other than supporting material of the structure and the incident wave will pass through it. So in order to verify this we performed the experimental simulation and verification process on various thickness parameters of the substrate and visualized the response on the transmissive and the reflective properties. In the analysis we used copper a lossy material as element to make a double square loop structure.

Main advantage of using Double Square Loop is that it allows us to get bandpass behaviour and there is no effect of polarisation on the output characteristics.

Table 1. Dimension Parameters of the Designed Double Square Loop FSSs (f_1 and f_2 in GHz, L_1 and L_2 in nH, C_1 and C_2 in pF and all structural dimensions are in mm).

These were the structural parameters taken in our simulation to get the desired response. In the Table 1, d_1 , d_2 , w_1 and w_2 are being lengths and widths of inner and outer loop respectively. The algorithmic flow of our process is as:

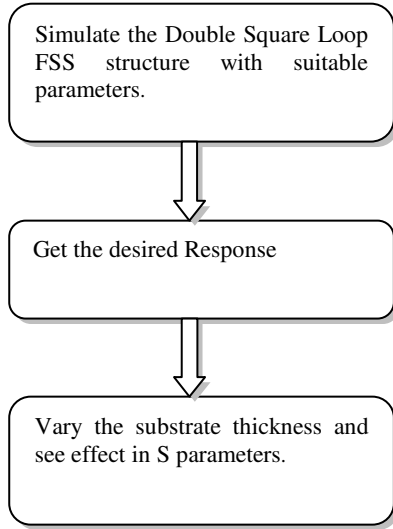


Fig. 3. Algorithmic Flow

4. EXPERIMENTAL RESULTS

We have simulated our structure in the Ka band of the microwave spectrum and got the desired response. In this paper we have studied the effect of varying the substrate thickness in the transmissive and the reflective properties of the structure. Double Square Loop structure was chosen and simulated with embedding it on the Thermocol substrate. The incident electromagnetic wave was chosen and being passed through it and the parameters were adjusted to get the response within the desired range. We have done analysis and got the responses which will help the researchers in designing their structures. We choose the dielectric normal material i.e. Thermocol with $\epsilon = 1.15$, $\mu = 1$ and dimensions same as that of periodicity (p).

A. Effect of Substrate with thickness 0.2mm

Substrate with dimensions from -0.2 to 0 was chosen and with overall thickness of 0.2 we visualized the bandpass behavior was excellent and the cut-off frequencies were also sharp.

So as seen in the figure the cutoff frequencies are sharp and almost ideal graph of the bandpass filter is visualized. The substrate plays a vital role in handling our FSS structure and making it feasible in the real world.

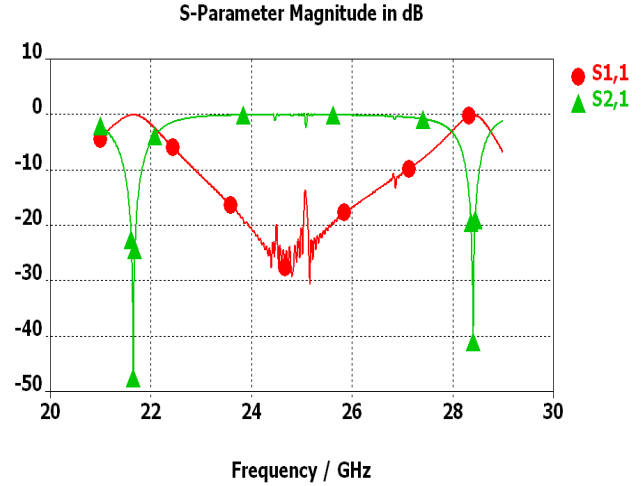


Fig. 4. Transmission and reflection Characteristics at thickness 0.2mm.

B. Effect of Substrate with thickness 2mm

When we increased the substrate thickness to 2mm we can visualize from the below result that Reflection parameter S_{11} varies sharply thus increasing the reflection around 25GHz.

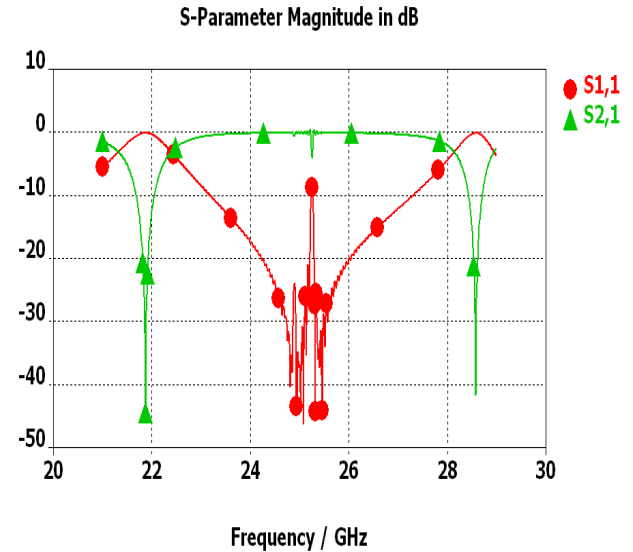


Fig. 5. Transmission and reflection Characteristics at thickness 2mm.

Hence it should be taken into account when analyzing the FSS structure.

C. Effect of substrate with thickness 4mm.

On further increasing the thickness of the substrate to 4mm we can see the losses tend to scatter in the whole microwave

spectrum taken into account making different notches of increasing transmission and reflection magnitudes.

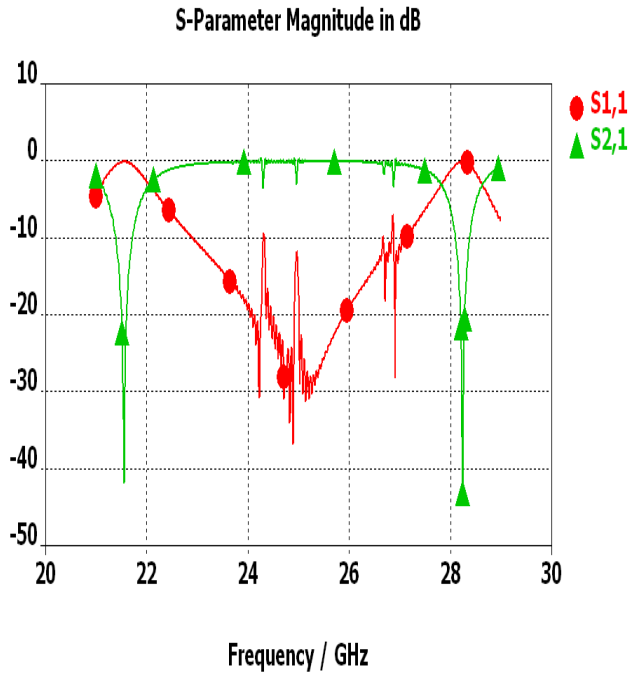


Fig. 6. Transmission and reflection Characteristics at thickness 4mm.

Note that the notches in the S_{11} parameter tends to increase and scatter in the whole band taken into account. We have analysed the Double Square Loop structure for the frequency range of 24-28GHz which will give us approximately 15% transmission band. The analysis done on the CST simulation software with consideration of incident angle to be 0° and we have chosen substrate of epsilon 1.15 and dimensions same as that of p on which our structure is placed.

5. CONCLUSIONS

Frequency Selective Surfaces are still an ongoing research field and this article will prove to be beneficial for the researchers to develop FSS with more accuracy. The results were visualized and it is seen that as we increase the thickness of the substrate the losses goes on increasing and the distortions tend to occur in our transmission and the reflection band. No doubt that it will not affect our cutoff frequencies but will increase distortions and reflections in the transmission band. In order to remove these discrepancies we must adjust and analyze the proper width of the substrate to continue our design process in the right direction. So, by studying the effect of the substrate thickness, the design of the Frequency Selective Surface may be adjusted to give the desired transmission and reflection characteristics.

REFERENCES

- [1] B. A. Munk, "Frequency Selective Surfaces: Theory and Design", New York: Wiley, 2000, pp. 2–23.
- [2] T. K. Wu, "Frequency Selective Surface and Grid Array", New York: Wiley, 1995.
- [3] R. Mittra, C. H. Tsao, and W. L. Ko, "Frequency selective surfaces with applications in microwaves and optics," in Proc. IEEE Microwave Symp, vol. 80, pp. 447–449, 1980.
- [4] Abbaspour-Tamijani, K. S., and G. M. Rebeiz, "Antenna Filter-Antenna Arrays as a Class of Band-Pass Frequency Selective Surfaces", IEEE Trans. Microw. Theory Tech. , vol 52 , pp . 1781 -1789, 2004.
- [5] K.Sarabandi and N. Behdad, "A Frequency Selective Surface with Miniaturized Elements," IEEE Trans. Antennas Propag. , vol 55, pp . 1239-1245, 2007 .
- [6] N. Marcuvitz, "Waveguide Handbook" (1st edition), New York: McGraw-Hill, 1951.
- [7] R.J. Langley, E.A. Parker, "Double square frequency selective surfaces and their equivalent circuit" , Electronics Letters, vol. 19, no. 17, pp. 675 – 677, 1983.
- [8] E.F. Kent, B. Doken and M. Kartal "A New Equivalent Circuit Based FSS Design Method by Using Genetic Algorithm", International Conference on Engineering Optimization , Lisbon, Portugal, Sept. 2010.
- [9] L. Zappelli, "Analysis of modified dielectric frequency selective surfaces under 3-D plane wave excitation using a multimode equivalent network approach," IEEE Trans. Antennas Propag. , Vol. 57, No. 4, pp. 1105–1114, April 2009..
- [10] F. Bayatpur, K. Sarabandi, "Single-Layer, High-Order, Miniaturized-Element Frequency Selective Surfaces," IEEE Transactions on Microwave Theory and Techniques, Nov. 2008.
- [11] Y.J Lee, J. Yeo, R. Mittra, and W. S. Park, "Application of electromagnetic bandgap (EBG) superstrates with controllable defects for a class of patch antennas as spatial angular filters", IEEE Trans. Antennas and Propag. , vol. AP-53, no. 1, pp. 224–234, Jan. 2005.
- [12] S. Biber, M. Bozzi, O. Gunther, L. Perregrini, and L. P. Schmidt, "De-sign and testing of frequency-selective surfaces on silicon substrates for submillimeter-wave applications," IEEE Trans. Antennas Propag., Vol. 54, No. 9, pp. 2638–2645, 2006.
- [13] J.P. Gianvittorio, J. Zendejas, Y. Rahmat-Samii and J. Judy, "Reconfigurable MEMS-enabled frequency selective surfaces," IEE Electron.Lett., Vol. 38, No. 25, pp. 1627–1628, Dec. 2002.
- [14] M. Ohira, H. Deguchi, M. Tsuji, and H. Shigesawa, "Multiband single-layer frequency selective surface designed by combination of genetic algorithm and geometry-refinement technique," IEEE Trans. Antennas Propag., vol. 52, no. 11, pp. 2925–2931, Nov. 2004.
- [15] R.A.Hill and B.A. Munk, "The effect of perturbing a frequency selective surface and its relation to the design of a dual-band surface," IEEE Trans. Antennas Propagat, vol. AP-44, no.3, pp.368-374, Mar. 1996.
- [16] A.E. Yilmaz and M. Kuzuoglu, "Design of the square loop frequency selective surfaces with particle swarm optimization via the equivalent circuit model, vol. 18, no. 2, pp. 95-102, 2009.

PARTNERS



सत्यमेव जयते

**Ministry of
Communications and
Information Technology**



TECHNOLOGY PARTNER



International Journal of Advances in
Electrical and Electronics Engineering (IJAEEE)



Excellent Publishing House

Kishangarh, Vasant Kunj, New Delhi-110 070

Contact : 9910948516, 9958167102

e-mail : exlpublishers@gmail.com

ISBN: 978-93-81583-93-7



9 789381 583937