IJCSI

# International Journal of Computer Science Issues

Volume 9, Issue 1, No 2, January 2012
ISSN (Online): 1694-0814

**IJCSI proceedings are currently indexed by:**

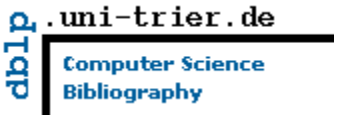
















# IJCSI Publicity Board 2012

# IJCSI Editorial Board 2012

# EDITORIAL

In this first edition of 2012, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

It was with pleasure and a sense of satisfaction that we announced in mid March 2011 our 2-year Impact Factor which is evaluated at 0.242. For more information about this please see the 3rd question in FAQ section of the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 9, Issue 1, No 2, January 2012 (IJCSI Vol. 9, Issue 1, No 2). The acceptance rate for this issue is 30.1%.

IJCSI Editorial Board
January 2012 Issue
ISSN (Online): 1694-0814

www.IJCSI.org

# IJCSI Reviewers Committee 2012

- Mr. Markus Schatten, University of Zagreb, Faculty of Organization and Informatics, Croatia
- Mr. Vassilis Papataxiarhis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece
- Dr Modestos Stavrakis, University of the Aegean, Greece
- Dr Fadi KHALIL, LAAS -- CNRS Laboratory, France
- Dr Dimitar Trajanov, Faculty of Electrical Engineering and Information technologies, ss. Cyril and Methodius Univesity - Skopje, Macedonia
- Dr Jinping Yuan, College of Information System and Management,National Univ. of Defense Tech., China
- Dr Alexis Lazanas, Ministry of Education, Greece
- Dr Stavroula Mougiakakou, University of Bern, ARTORG Center for Biomedical Engineering Research, Switzerland
- Dr Cyril de Runz, CReSTIC-SIC, IUT de Reims, University of Reims, France
- Mr. Pramodkumar P. Gupta, Dept of Bioinformatics, Dr D Y Patil University, India
- Dr Alireza Fereidunian, School of ECE, University of Tehran, Iran
- Mr. Fred Viezens, Otto-Von-Guericke-University Magdeburg, Germany
- Dr. Richard G. Bush, Lawrence Technological University, United States
- Dr. Ola Osunkoya, Information Security Architect, USA
- Mr. Kotsokostas N.Antonios, TEI Piraeus, Hellas
- Prof Steven Totosy de Zepetnek, U of Halle-Wittenberg & Purdue U & National Sun Yat-sen U, Germany, USA, Taiwan
- Mr. M Arif Siddiqui, Najran University, Saudi Arabia
- Ms. Ilknur Icke, The Graduate Center, City University of New York, USA
- Prof Miroslav Baca, Faculty of Organization and Informatics, University of Zagreb, Croatia
- Dr. Elvia Ruiz Beltrán, Instituto Tecnológico de Aguascalientes, Mexico
- Mr. Moustafa Banbouk, Engineer du Telecom, UAE
- Mr. Kevin P. Monaghan, Wayne State University, Detroit, Michigan, USA
- Ms. Moira Stephens, University of Sydney, Australia
- Ms. Maryam Feily, National Advanced IPv6 Centre of Excellence (NAV6) , Universiti Sains Malaysia (USM), Malaysia
- Dr. Constantine YIALOURIS, Informatics Laboratory Agricultural University of Athens, Greece
- Mrs. Angeles Abella, U. de Montreal, Canada
- Dr. Patrizio Arrigo, CNR ISMAC, italy
- Mr. Anirban Mukhopadhyay, B.P.Poddar Institute of Management & Technology, India
- Mr. Dinesh Kumar, DAV Institute of Engineering & Technology, India
- Mr. Jorge L. Hernandez-Ardieta, INDRA SISTEMAS / University Carlos III of Madrid, Spain
- Mr. AliReza Shahrestani, University of Malaya (UM), National Advanced IPv6 Centre of Excellence (NAv6), Malaysia
- Mr. Blagoj Ristevski, Faculty of Administration and Information Systems Management - Bitola, Republic of Macedonia
- Mr. Mauricio Egidio Cantão, Department of Computer Science / University of São Paulo, Brazil
- Mr. Jules Ruis, Fractal Consultancy, The netherlands

- Mr. Mohammad Iftekhar Husain, University at Buffalo, USA
- Dr. Deepak Laxmi Narasimha, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
- Dr. Paola Di Maio, DMEM University of Strathclyde, UK
- Dr. Bhanu Pratap Singh, Institute of Instrumentation Engineering, Kurukshetra University Kurukshetra, India
- Mr. Sana Ullah, Inha University, South Korea
- Mr. Cornelis Pieter Pieters, Condast, The Netherlands
- Dr. Amogh Kavimandan, The MathWorks Inc., USA
- Dr. Zhinan Zhou, Samsung Telecommunications America, USA
- Mr. Alberto de Santos Sierra, Universidad Politécnica de Madrid, Spain
- Dr. Md. Atiqur Rahman Ahad, Department of Applied Physics, Electronics & Communication Engineering (APECE), University of Dhaka, Bangladesh
- Dr. Charalampos Bratsas, Lab of Medical Informatics, Medical Faculty, Aristotle University, Thessaloniki, Greece
- Ms. Alexia Dini Kounoudes, Cyprus University of Technology, Cyprus
- Dr. Jorge A. Ruiz-Vanoye, Universidad Juárez Autónoma de Tabasco, Mexico
- Dr. Alejandro Fuentes Penna, Universidad Popular Autónoma del Estado de Puebla, México
- Dr. Ocotlán Díaz-Parra, Universidad Juárez Autónoma de Tabasco, México
- Mrs. Nantia Iakovidou, Aristotle University of Thessaloniki, Greece
- Mr. Vinay Chopra, DAV Institute of Engineering & Technology, Jalandhar
- Ms. Carmen Lastres, Universidad Politécnica de Madrid - Centre for Smart Environments, Spain
- Dr. Sanja Lazarova-Molnar, United Arab Emirates University, UAE
- Mr. Srikrishna Nudurumati, Imaging & Printing Group R&D Hub, Hewlett-Packard, India
- Dr. Olivier Nocent, CReSTIC/SIC, University of Reims, France
- Mr. Burak Cizmeci, Isik University, Turkey
- Dr. Carlos Jaime Barrios Hernandez, LIG (Laboratory Of Informatics of Grenoble), France
- Mr. Md. Rabiul Islam, Rajshahi university of Engineering & Technology (RUET), Bangladesh
- Dr. LAKHOUA Mohamed Najeh, ISSAT - Laboratory of Analysis and Control of Systems, Tunisia
- Dr. Alessandro Lavacchi, Department of Chemistry - University of Firenze, Italy
- Mr. Mungwe, University of Oldenburg, Germany
- Mr. Somnath Tagore, Dr D Y Patil University, India
- Ms. Xueqin Wang, ATCS, USA
- Dr. Borislav D Dimitrov, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland
- Dr. Fondjo Fotou Franklin, Langston University, USA
- Dr. Vishal Goyal, Department of Computer Science, Punjabi University, Patiala, India
- Mr. Thomas J. Clancy, ACM, United States
- Dr. Ahmed Nabih Zaki Rashed, Dr. in Electronic Engineering, Faculty of Electronic Engineering, menouf 32951, Electronics and Electrical Communication Engineering Department, Menoufia university, EGYPT, EGYPT
- Dr. Rushed Kanawati, LIPN, France
- Mr. Koteshwar Rao, K G Reddy College Of ENGG.&TECH,CHILKUR, RR DIST.,AP, India
- Mr. M. Nagesh Kumar, Department of Electronics and Communication, J.S.S. research foundation, Mysore University, Mysore-6, India

- Dr. Ibrahim Noha, Grenoble Informatics Laboratory, France
- Mr. Muhammad Yasir Qadri, University of Essex, UK
- Mr. Annadurai .P, KMCPGS, Lawspet, Pondicherry, India, (Aff. Pondicherry Univeristy, India
- Mr. E Munivel , CEDTI (Govt. of India), India
- Dr. Chitra Ganesh Desai, University of Pune, India
- Mr. Syed, Analytical Services & Materials, Inc., USA
- Mrs. Payal N. Raj, Veer South Gujarat University, India
- Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal, India
- Mr. Mahesh Goyani, S.P. University, India, India
- Mr. Vinay Verma, Defence Avionics Research Establishment, DRDO, India
- Dr. George A. Papakostas, Democritus University of Thrace, Greece
- Mr. Abhijit Sanjiv Kulkarni, DARE, DRDO, India
- Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
- Dr. B. Sivaselvan, Indian Institute of Information Technology, Design & Manufacturing, Kancheepuram, IIT Madras Campus, India
- Dr. Partha Pratim Bhattacharya, Greater Kolkata College of Engineering and Management, West Bengal University of Technology, India
- Mr. Manish Maheshwari, Makhanlal C University of Journalism & Communication, India
- Dr. Siddhartha Kumar Khaitan, Iowa State University, USA
- Dr. Mandhapati Raju, General Motors Inc, USA
- Dr. M.Iqbal Saripan, Universiti Putra Malaysia, Malaysia
- Mr. Ahmad Shukri Mohd Noor, University Malaysia Terengganu, Malaysia
- Mr. Selvakuberan K, TATA Consultancy Services, India
- Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
- Mr. Rakesh Kachroo, Tata Consultancy Services, India
- Mr. Raman Kumar, National Institute of Technology, Jalandhar, Punjab., India
- Mr. Nitesh Sureja, S.P.University, India
- Dr. M. Emre Celebi, Louisiana State University, Shreveport, USA
- Dr. Aung Kyaw Oo, Defence Services Academy, Myanmar
- Mr. Sanjay P. Patel, Sankalchand Patel College of Engineering, Visnagar, Gujarat, India
- Dr. Pascal Fallavollita, Queens University, Canada
- Mr. Jitendra Agrawal, Rajiv Gandhi Technological University, Bhopal, MP, India
- Mr. Ismael Rafael Ponce Medellín, Cenidet (Centro Nacional de Investigación y Desarrollo Tecnológico), Mexico
- Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan
- Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe
- Mrs. Mutalli Vatila, Offshore Business Philipines, Philipines
- Mr. Pankaj Kumar, SAMA, India
- Dr. Himanshu Aggarwal, Punjabi University,Patiala, India
- Dr. Vauvert Guillaume, Europages, France
- Prof Yee Ming Chen, Department of Industrial Engineering and Management, Yuan Ze University, Taiwan
- Dr. Constantino Malagón, Nebrija University, Spain
- Prof Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
- Mr. Angkoon Phinyomark, Prince of Singkla University, Thailand

- Ms. Nital H. Mistry, Veer Narmad South Gujarat University, Surat, India
- Dr. M.R.Sumalatha, Anna University, India
- Mr. Somesh Kumar Dewangan, Disha Institute of Management and Technology, India
- Mr. Raman Maini, Punjabi University, Patiala(Punjab)-147002, India
- Dr. Abdelkader Outtagarts, Alcatel-Lucent Bell-Labs, France
- Prof Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
- Mr. Prabu Mohandas, Anna University/Adhiyamaan College of Engineering, india
- Dr. Manish Kumar Jindal, Panjab University Regional Centre, Muktsar, India
- Prof Mydhili K Nair, M S Ramaiah Institute of Technnology, Bangalore, India
- Dr. C. Suresh Gnana Dhas, VelTech MultiTech Dr.Rangarajan Dr.Sagunthala Engineering College,Chennai,Tamilnadu, India
- Prof Akash Rajak, Krishna Institute of Engineering and Technology, Ghaziabad, India
- Mr. Ajay Kumar Shrivastava, Krishna Institute of Engineering & Technology, Ghaziabad, India
- Dr. Vu Thanh Nguyen, University of Information Technology HoChiMinh City, VietNam
- Prof Deo Prakash, SMVD University (A Technical University open on I.I.T. Pattern) Kakryal (J&K), India
- Dr. Navneet Agrawal, Dept. of ECE, College of Technology & Engineering, MPUAT, Udaipur 313001 Rajasthan, India
- Mr. Sufal Das, Sikkim Manipal Institute of Technology, India
- Mr. Anil Kumar, Sikkim Manipal Institute of Technology, India
- Dr. B. Prasanalakshmi, King Saud University, Saudi Arabia.
- Dr. K D Verma, S.V. (P.G.) College, Aligarh, India
- Mr. Mohd Nazri Ismail, System and Networking Department, University of Kuala Lumpur (UniKL), Malaysia
- Dr. Nguyen Tuan Dang, University of Information Technology, Vietnam National University Ho Chi Minh city, Vietnam
- Dr. Abdul Aziz, University of Central Punjab, Pakistan
- Dr. P. Vasudeva Reddy, Andhra University, India
- Mrs. Savvas A. Chatzichristofis, Democritus University of Thrace, Greece
- Mr. Marcio Dorn, Federal University of Rio Grande do Sul - UFRGS Institute of Informatics, Brazil
- Mr. Luca Mazzola, University of Lugano, Switzerland
- Mr. Hafeez Ullah Amin, Kohat University of Science & Technology, Pakistan
- Dr. Professor Vikram Singh, Ch. Devi Lal University, Sirsa (Haryana), India
- Dr. Shahanawaj Ahamad, Department of Computer Science, King Saud University, Saudi Arabia
- Dr. K. Duraiswamy, K. S. Rangasamy College of Technology, India
- Prof. Dr Mazlina Esa, Universiti Teknologi Malaysia, Malaysia
- Dr. P. Vasant, Power Control Optimization (Global), Malaysia
- Dr. Taner Tuncer, Firat University, Turkey
- Dr. Norrozila Sulaiman, University Malaysia Pahang, Malaysia
- Prof. S K Gupta, BCET, Guradspur, India
- Dr. Latha Parameswaran, Amrita Vishwa Vidyapeetham, India
- Mr. M. Azath, Anna University, India
- Dr. P. Suresh Varma, Adikavi Nannaya University, India
- Prof. V. N. Kamalesh, JSS Academy of Technical Education, India
- Dr. D Gunaseelan, Ibri College of Technology, Oman

- Mr. Sanjay Kumar Anand, CDAC, India
- Mr. Akshat Verma, CDAC, India
- Mrs. Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
- Mr. Hasan Asil, Islamic Azad University Tabriz Branch (Azarshahr), Iran
- Prof. Dr Sajal Kabiraj, Fr. C Rodrigues Institute of Management Studies (Affiliated to University of Mumbai, India), India
- Mr. Syed Fawad Mustafa, GAC Center, Shandong University, China
- Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
- Prof. Selvakani Kandeeban, Francis Xavier Engineering College, India
- Mr. Tohid Sedghi, Urmia University, Iran
- Dr. S. Sasikumar, PSNA College of Engg and Tech, Dindigul, India
- Dr. Anupam Shukla, Indian Institute of Information Technology and Management Gwalior, India
- Mr. Rahul Kala, Indian Institute of Inforamtion Technology and Management Gwalior, India
- Dr. A V Nikolov, National University of Lesotho, Lesotho
- Mr. Kamal Sarkar, Department of Computer Science and Engineering, Jadavpur University, India
- Prof. Sattar J Aboud, Iraqi Council of Representatives, Iraq-Baghdad
- Dr. Prasant Kumar Pattnaik, Department of CSE, KIST, India
- Dr. Mohammed Amoon, King Saud University, Saudi Arabia
- Dr. Tsvetanka Georgieva, Department of Information Technologies, St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
- Mr. Ujjal Marjit, University of Kalyani, West-Bengal, India
- Dr. Prasant Kumar Pattnaik, KIST,Bhubaneswar,India, India
- Dr. Guezouri Mustapha, Department of Electronics, Faculty of Electrical Engineering, University of Science and Technology (USTO), Oran, Algeria
- Mr. Maniyar Shiraz Ahmed, Najran University, Najran, Saudi Arabia
- Dr. Sreedhar Reddy, JNTU, SSIETW, Hyderabad, India
- Mr. Bala Dhandayuthapani Veerasamy, Mekelle University, Ethiopa
- Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
- Mr. Rajesh Prasad, LDC Institute of Technical Studies, Allahabad, India
- Ms. Habib Izadkhah, Tabriz University, Iran
- Dr. Lokesh Kumar Sharma, Chhattisgarh Swami Vivekanand Technical University Bhilai, India
- Mr. Kuldeep Yadav, IIIT Delhi, India
- Dr. Naoufel Kraiem, Institut Superieur d'Informatique, Tunisia
- Prof. Frank Ortmeier, Otto-von-Guericke-Universitaet Magdeburg, Germany
- Mr. Ashraf Aljammal, USM, Malaysia
- Mrs. Amandeep Kaur, Department of Computer Science, Punjabi University, Patiala, Punjab, India
- Mr. Babak Basharirad, University Technology of Malaysia, Malaysia
- Mr. Avinash singh, Kiet Ghaziabad, India
- Dr. Miguel Vargas-Lombardo, Technological University of Panama, Panama
- Dr. Tuncay Sevindik, Firat University, Turkey
- Ms. Pavai Kandavelu, Anna University Chennai, India
- Mr. Ravish Khichar, Global Institute of Technology, India
- Mr Aos Alaa Zaidan Ansaef, Multimedia University, Cyberjaya, Malaysia
- Dr. Awadhesh Kumar Sharma, Dept. of CSE, MMM Engg College, Gorakhpur-273010, UP, India
- Mr. Qasim Siddique, FUIEMS, Pakistan

- Dr. Le Hoang Thai, University of Science, Vietnam National University - Ho Chi Minh City, Vietnam
- Dr. Saravanan C, NIT, Durgapur, India
- Dr. Vijay Kumar Mago, DAV College, Jalandhar, India
- Dr. Do Van Nhon, University of Information Technology, Vietnam
- Dr. Georgios Kioumourtzis, Researcher, University of Patras, Greece
- Mr. Amol D.Potgantwar, SITRC Nasik, India
- Mr. Lesedi Melton Masisi, Council for Scientific and Industrial Research, South Africa
- Dr. Karthik.S, Department of Computer Science & Engineering, SNS College of Technology, India
- Mr. Nafiz Imtiaz Bin Hamid, Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Bangladesh
- Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
- Dr. Abdul Kareem M. Radhi, Information Engineering - Nahrin University, Iraq
- Dr. Manuj Darbari, BBDNITM, Institute of Technology, A-649, Indira Nagar, Lucknow 226016, India
- Ms. Izerrouken, INP-IRIT, France
- Mr. Nitin Ashokrao Naik, Dept. of Computer Science, Yeshwant Mahavidyalaya, Nanded, India
- Mr. Nikhil Raj, National Institute of Technology, Kurukshetra, India
- Prof. Maher Ben Jemaa, National School of Engineers of Sfax, Tunisia
- Prof. Rajeshwar Singh, BRCM College of Engineering and Technology, Bahal Bhiwani, Haryana, India
- Mr. Gaurav Kumar, Department of Computer Applications, Chitkara Institute of Engineering and Technology, Rajpura, Punjab, India
- Mr. Ajeet Kumar Pandey, Indian Institute of Technology, Kharagpur, India
- Mr. Rajiv Phougat, IBM Corporation, USA
- Mrs. Aysha V, College of Applied Science Pattuvam affiliated with Kannur University, India
- Dr. Debotosh Bhattacharjee, Department of Computer Science and Engineering, Jadavpur University, Kolkata-700032, India
- Dr. Neelam Srivastava, Institute of engineering & Technology, Lucknow, India
- Prof. Sweta Verma, Galgotia's College of Engineering & Technology, Greater Noida, India
- Mr. Harminder Singh BIndra, MIMIT, INDIA
- Mr. Tarun Kumar, U.P. Technical University/Radha Govinend Engg. College, India
- Mr. Tirthraj Rai, Jawahar Lal Nehru University, New Delhi, India
- Mr. Akhilesh Tiwari, Madhav Institute of Technology & Science, India
- Mr. Dakshina Ranjan Kisku, Dr. B. C. Roy Engineering College, WBUT, India
- Ms. Anu Suneja, Maharshi Markandeshwar University, Mullana, Haryana, India
- Mr. Munish Kumar Jindal, Punjabi University Regional Centre, Jaito (Faridkot), India
- Dr. Ashraf Bany Mohammed, Management Information Systems Department, Faculty of Administrative and Financial Sciences, Petra University, Jordan
- Mrs. Jyoti Jain, R.G.P.V. Bhopal, India
- Dr. Lamia Chaari, SFAX University, Tunisia
- Mr. Akhter Raza Syed, Department of Computer Science, University of Karachi, Pakistan
- Prof. Khubaib Ahmed Qureshi, Information Technology Department, HIMS, Hamdard University, Pakistan
- Prof. Boubker Sbihi, Ecole des Sciences de L'Information, Morocco
- Dr. S. M. Riazul Islam, Inha University, South Korea
- Prof. Lokhande S.N., S.R.T.M.University, Nanded (MH), India
- Dr. Vijay H Mankar, Dept. of Electronics, Govt. Polytechnic, Nagpur, India

- Mr. Ojesanmi Olusegun, Ajayi Crowther University, Oyo, Nigeria
- Ms. Mamta Juneja, RBIEBT, PTU, India
- Prof. Chandra Mohan, John Bosco Engineering College, India
- Dr. Bodhe Shrikant K., College of Engineering, Pandhapur, Maharashtra, INDIA
- Dr. Sherif G. Aly, The American University in Cairo, Egypt
- Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India
- Prof. Nikhil gondaliya, G H Patel College of Engg. & Technology, India
- Mr. Nisheeth Joshi, Apaji Institute, Banasthali University, India
- Mr. Nizar, National Ingineering School of Monastir, Tunisia
- Prof. R. Jagadeesh Kannan, RMK Engineering College, India
- Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India
- Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India
- Dr. Intisar A. M. Al Sayed, Associate prof./College of Science and IT/Al Isra University, Jordan
- Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India
- Mrs. Rajalakshmi, JIITU, India
- Mr. Shrikant Ardhapurkar, Indian Institute of Information Techonology, India
- Ms. Hemalatha R, Osmania University, India
- Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia
- Mr. Sunil Kumar Grandhi, Maris Stella College, India
- Prof. Shishir K. Shandilya, NRI Institute of Science & Technology, INDIA
- Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
- Prof. Prasun Ghosal, Bengal Engineering and Science University, India
- Dr. Nagarajan Velmurugan, SMVEC/Pondicherry University, India
- Dr. R. Baskaran, Anna University, India
- Dr. Wichian Sittiprapaporn, Mahasarakham University College of Music, Thailand
- Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
- Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology, India
- Mrs. Inderpreet Kaur, PTU, Jalandhar, India
- Mr. Palaniyappan, K7 Virus Research Laboratory, India
- Mr. Guanbo Zheng, University of Houston, main campus, USA
- Mr. Arun Kumar Tripathi, Krishna Institute of Engg. and Tech-Ghaziabad, Affilated to UPTU, India
- Mr. Iqbaldeep Kaur, PTU / RBIEBT, India
- Mr. Amit Choudhary, Maharaja Surajmal Institute, New Delhi, India
- Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India
- Dr. Ashish Avasthi, Uttar Pradesh Technical University, India
- Dr. Manish Kumar, Uttar Pradesh Technical University, India
- Prof. Vinay Uttamrao Kale, P.R.M. Institute of Technology & Research, Badnera, Amravati, Maharashtra, India
- Mr. Suhas J Manangi, Microsoft, India
- Mr. Shyamalendu Kandar, Haldia Institute of Technology, India
- Ms. Anna Kuzio, Adam Mickiewicz University, School of English, Poland
- Mr. Vikas Singla, Malout Institute of Management & Information Technology, Malout, Punjab, India, India

- Dr. Dalbir Singh, Faculty of Information Science And Technology, National University of Malaysia, Malaysia
- Dr. Saurabh Mukherjee, PIM, Jiwaji University, Gwalior, M.P, India
- Mr. Senthilnathan T, Sri Krishna College of Engineering and Technology, India
- Dr. Debojyoti Mitra, Sir Padampat Singhania University, India
- Prof. Rachit Garg, Department of Computer Science, L K College, India
- Dr. Arun Kumar Gupta, M.S. College, Saharanpur, India
- Dr. Todor Todorov, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria
- Mrs. Manjula K A, Kannur University, India
- Mrs. Sasikala R., K S R College of Technology, India
- Prof. M. Saleem Babu, Department of Computer Science and Engineering, Vel Tech University, Chennai, India
- Dr. Rajesh Kumar Tiwari, GLA Institute of Technology, India
- Mr. Rakesh Kumar, Indian Institute of Technology Roorkee, India
- Prof. Amit Verma, PTU/RBIEBT, India
- Mr. Sohan Purohit, University of Massachusetts Lowell, USA
- Mr. Anand Kumar, AMC Engineering College, Bangalore, India
- Dr. Samir Abdelrahman, Computer Science Department, Cairo University, Egypt
- Dr. Rama Prasad V Vaddella, Sree Vidyanikethan Engineering College, India
- Dr. Manoj Wadhwa, Echelon Institute of Technology Faridabad, India
- Mr. Zeashan Hameed Khan, UniversitÃ© de Grenoble, France
- Mr. Arup Kumar Pal, Indian SChool of Mines, Dhanbad, India
- Dr. Pouya, Islamic Azad University,Naein Branch, iran
- Prof. Jyoti Prakash Singh, Academy of Technology, India
- Mr. Muraleedharan CV, Sree Chitra Tirunal Institute for Medical Sciences & Technology, India
- Dr. E U Okike, University of Ibadan, Nigeria Kampala Int Univ Uganda, Nigeria
- Dr. D. S. Rao, Chitkara University, India
- Mr. Peyman Taher, Oklahoma State University, USA
- Dr. S Srinivasan, PDM College of Engineering, India
- Dr. Rafiqul Zaman Khan, Department of Computer Science, AMU, Aligarh, India
- Ms. Meenakshi Kalia, Shobhit University, India
- Mr. Muhammad Zakarya, Abdul Wali Khan University, Mardan, Pakistan, Pakistan
- Dr. M Gobi, PSG college, India
- Mr. Williamjeet Singh, Chitkara Institute of Engineering and Technology, India
- Mr. G.Jeyakumar, Amrita School of Engineering, India
- Mr. Osama Sohaib, University of Balochistan, Pakistan
- Mr. Jude Hemanth, Karunya University, India
- Mr. Nitin Rakesh, Jaypee University of Information Technology, India
- Mr. Harmunish Taneja, Maharishi Markandeshwar University, Mullana, Ambala, Haryana, India
- Dr. Sin-Ban Ho, Faculty of IT, Multimedia University, Malaysia
- Dr. Mashiur Rahman, Institute for Molecular Science, Japan
- Mrs. Doreen Hephzibah Miriam, Anna University, Chennai, India
- Mr. Kosala Yapa Bandara, Dublin City University, Ireland.
- Mrs. Mitu Dhull, GNKITMS Yamuna Nagar Haryana, India

- Dr. Chitra A.Dhawale, Professor, Symbiosis Institute of Computer Studies and Research, Pune (MS), India
- Dr. Arun Sharma, GB Technical University, Noida, India
- Mr. Naoufel Machta, Faculty of Science of Tunis, Tunisia
- Dr. Utpal Biswas, University of Kalyani, India
- Prof. Parma Nand, IIT Roorkee, India
- Prof. Mahesh P K, Jnana Vikas Institute of Tevhnology, Bangalore, India
- Dr. D.I. George Amalarethinam, Jamal Mohamed College, Bharathidasan University, India
- Mr. Ishtiaq ahmad, University of Engineering & Technology, Taxila, Pakistan
- Mrs. B.Sharmila, Sri Ramakrishna Engineering College, Coimbatore Anna University Coimbatore, India
- Dr. Muhammad Wasif Nisar, COMSATS Institue of Information Technology, Pakistan
- Mr. Prabu Dorairaj, EMC Corporation, India/USA
- Mr. Neetesh Gupta, Technocrats Inst. of Technology, Bhopal, India
- Dr. Ola Osunkoya, PRGX, USA
- Ms. A. Lavanya, Manipal University, Karnataka, India
- Dr. Jalal Laassiri, MIA-Laboratory, Faculty of Sciences Rabat, Morocco
- Mr. Ganesan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India
- Mr. V.Ramakrishnan, Sri Venkateswara college of Engineering and Technology, Thiruvallur, India
- Prof. Vuda Sreenivasarao, St. Mary's college of Engg & Tech, India
- Prof. Ashutosh Kumar Dubey, Assistant Professor, India
- Dr. R.Ramesh, Anna University, India
- Mr. Ali Khadair HMood, University of Malaya, Malaysia
- Dr. Vimal Mishra, U.P. Technical Education, India
- Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India
- Mrs. D.Suganyadevi, SNR SONS College (Autonomous), India
- Mr. Prasad S.Halgaonkar, MIT, Pune University, India
- Mr. Vijay Kumar, College of Engg. and Technology, IFTM, Moradabad(U.P), India
- Mr. Mehran Parchebafieh, Douran, Iran
- Mr. Anand Sharma, MITS, Lakshmangarh, Sikar (Rajasthan), India
- Mr. Amit Kumar, Jaypee University of Engineering and Technology, India
- Prof. B.L.Shivakumar, SNR Sons College, Coimbatore, India
- Mr. Mohammed Imran, JMI, India
- Dr. R Bremananth, School of EEE, Information Engineering (Div.), Nanyang Technological University, Singapore
- Prof. Vasavi Bande, Computer Science and Engneering, Hyderabad Institute of Technology and Management, India
- Dr. S.R.Balasundaram, National Institute of Technology, India
- Dr. Prasart Nuangchalerm, Mahasarakham University, Thailand
- Dr. M Ayoub Khan, C-DAC, Ministry of Communications & IT., India
- Dr. Jagdish Lal Raheja, Central Electronics Engineering Research Institute, India
- Mr G. Appasami, Dept. of CSE, Dr. Pauls Engineering College, Anna University - Chennai, India
- Mr Vimal Mishra, U.P. Technical Education, Allahabad, India
- Mr. Amin Daneshmand Malayeri, Young Researchers Club, Islamic AZAD University, Malayer Branch, Iran
- Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India

- Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India
- Dr. Soumen Mukherjee, RCC Institute of Information Technology, India
- Dr. Hamid Mcheick, University of Qubec at Chicoutimi, Canada
- Dr. Mokhled AlTarawneh, PhD computer engineering/ Faulty of engineerin/ mutah university, jordan
- Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India
- Ms. Suman Lata, Rayat Bahara institue of engg. & Nanotechnology,Hoshiarpur, India
- Dr. Shaikh Abdul Hannan, Vivekanand College, Aurangabad, India
- Prof. PN Kumar, Amrita Vishwa Vidyapeetham, India
- Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India
- Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh
- Mr. Sunil, Vignan College, India
- Mr. Ajit Singh, TIT&S Bhiwani, Haryana, India
- Mr. Nasim Qaisar, Federal Urdu Univetrsity of Arts , Science and Technology, Pakistan
- Ms. Rshma, Maharishi Markandeshwar University, India
- Mr. Gaurav Kumar Leekha, M.M.University, Solan (Himachal Pardesh), India
- Mr. Ordinor Tucker, Ministry of Finance Jamaica, Jamaica
- Mr. Mohit Jain, Maharaja Surajmal Institute of Technology (Affiliated to Guru Gobind Singh Indraprastha University, New Delhi), India
- Dr. Shaveta Rani, GZS College of Engineering & Technology, India
- Dr. Paramjeet Singh, GZS College of Engineering & Technology, India
- Dr. G R Sinha, SSCET, India
- Mr. Chetan Sharma, TechMahindra India Ltd., India
- Dr. Nabil Mohammed Ali Munassar, University of Science and Technology, Yemen
- Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management , India
- Prof. Vasavi Bande, HITAM, Engineering College, India
- Prof. S.P.Setty, Andhra University, India
- Dr. C. Kiran Mai, J.N.T.University,Hyderabad/VNR Vignana Jyothi Institute of Engineering & Technology/, India
- Ms. Bindiya Ahuja, Manav Rachna International University, India
- Mrs. Deepa Bura, Manav Rachna International University, India
- Mr. Vikas Gupta, CDLM Government Engineering College, Panniwala Mota, India
- Dr Juan José Martínez Castillo, University of Yacambu, Venezuela
- Mr Kunwar S. Vaisla, Department of Computer Science & Engineering, BCT Kumaon Engineering College, India
- Mr. Abhishek Shukla, RKGIT, India
- Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India
- Mr. Syed Imran, University College Cork, Ireland
- Dr. Intisar Al Said, Associate Prof/Al Isra University, Jordan
- Dr. Namfon Assawamekin, University of the Thai Chamber of Commerce, Thailand
- Dr. Shiv KUmar, Technocrat Institute of Technology-Bhopal (M.P.), India
- Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
- Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
- Dr. Mohamed Ali Mahjoub, University of Monastir, Tunisia

- Mr. Adis Medic, Infosys ltd, Bosnia and Herzegovina
- Mr Swarup Roy, Department of Information Technology, North Eastern Hill University, Umshing, Shillong 793022, Meghalaya, India
- Prof. Jakimi, Faculty of Science and technology my ismail University, Morocco
- Dr. R. Manicka Chezian, N G M College, Pollachi - 642 001, Tamilnadu, India
- Dr. P.Dananjayan, Pondicherry Engineering College, India
- Mr. Manik Sharma, Sewa Devi SD College Tarn Taran, India
- Mr. Suresh Kallam, East China University of Technology, Nanchang, China
- Dr. Mohammed Ali Hussain, Sai Madhavi Institute of Science & Technology, Rajahmundry, India
- Mr. Vikas Gupta, Adesh Instutute of Engineering & Technology, India
- Dr. Anuraag Awasthi, JV Womens University, Jaipur, India
- Dr. Mathura Prasad Thapliyal, Department of Computer Science, HNB Garhwal University (Centr al University), Srinagar (Garhwal), India
- Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia, Malaysia
- Mr. Adnan Qureshi, University of Jinan, Shandong, P.R.China, P.R.China
- Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
- Mr. Mueen Uddin, Universiti Teknologi Malaysia, Malaysia
- Mr. Manoj Gupta, Apex Institute of Engineering & Technology,Jaipur ( Affiliated to Rajasthan Technical University,Rajasthan), Indian
- Mr. S. Albert Alexander, Kongu Engineering College, India
- Dr. Shaidah Jusoh, Zarqa Private University, Jordan
- Dr. Dushmanta Mallick, KMBB College of Engineering and Technology, India
- Mr. Santhosh Krishna B.V, Hindustan University, India
- Dr. Tariq Ahamad Ahanger, Kausar College Of Computer Sciences, India
- Dr. Chi Lin, Dalian University of Technology, China
- Prof. VIJENDRA BABU.D, ECE Department, Aarupadai Veedu Institute of Technology, Vinayaka Missions University, India
- Mr. Raj Gaurang Tiwari, Gautam Budh Technical University, India
- Mrs. Jeysree J, SRM University, India
- Dr. C S Reddy, VIT University, India
- Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Bio-Technology, Kharar, India
- Mr. Muhammad Shuaib Qureshi, Iqra National University, Peshawar, Pakistan, Pakistan
- Dr Pranam Paul, Narula Institute of Technology Agarpara. Kolkata: 700109; West Bengal, India
- Dr. G. M. Nasira, Sasurie College of Enginering, (Affliated to Anna University of Technology Coimbatore), India
- Dr. Manasawee Kaenampornpan, Mahasarakham University, Thailand
- Mrs. Iti Mathur, Banasthali University, India
- Mr. Avanish Kumar Singh, RRIMT, NH-24, B.K.T., Lucknow, U.P., India
- Mr. Velayutham Pavanasam, Adhiparasakthi Engineering College, Melmaruvathur, India
- Dr. Panagiotis Michailidis, University of Western Macedonia, Greece
- Mr. Amir Seyed Danesh, University of Malaya, Malaysia
- Dr. Nadeem Mahmood, Department of computer science, university of Karachi, Pakistan
- Dr. Terry Walcott, E-Promag Consultancy Group, United Kingdom
- Mr. Farhat Amine, High Institute of Management of Tunis, Tunisia
- Mr. Ali Waqar Azim, COMSATS Institute of Information Technology, Pakistan

- Mr. Zeeshan Qamar, COMSATS Institute of Information Technology, Pakistan
- Dr. Samsudin Wahab, MARA University of Technology, Malaysia
- Mr. Ashikali M. Hasan, CelNet Security, India
- Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT), India
- Mr. B V A N S S Prabhakar Rao, Dept. of CSE, Miracle Educational Society Group of Institutions, Vizianagaram, India
- Dr. T. Abdul Razak, Associate Professor of Computer Science, Jamal Mohamed College (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli-620020, India
- Mr. Aurobindo Ogra, University of Johannesburg, South Africa
- Mr. Essam Halim Houssein, Dept of CS - Faculty of Computers and Informatics, Benha - Egypt
- Dr. Hanumanthappa. J, DoS in Computer Science, India
- Mr. Rachit Mohan Garg, Jaypee University of Information Technology, India
- Mr. Kamal Kad, Infosys Technologies, Australia
- Mrs. Aditi Chawla, GNIT Group of Institutes, India
- Dr. Kumardatt Ganrje, Pune University, India
- Mr. Merugu Gopichand, JNTU/BVRIT, India
- Mr. Rakesh Kumar, M.M. University, Mullana,Ambala, India
- Mr. M. Sundar, IBM, India
- Prof. Mayank Singh, J.P. Institute of Engineering & Technology, India
- Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
- Mr. Khaleel Ahmad, S.V.S. University, India
- Mr. Amin Zehtabian, Babol Noshirvani University of Technology / Tetta Electronic Company, Iran
- Mr. Rahul Katarya, Department of Information Technology , Delhi Technological University, India
- Dr. Vincent Ele Asor, University of Port Harcourt, Nigeria
- Ms. Prayas Kad, Capgemini Australia Ltd, Australia
- Mr. Alireza Jolfaei, Faculty and Research Center of Communication and Information Technology, IHU, Iran
- Mr. Nitish Gupta, GGSIPU, India
- Dr. Mohd Lazim Abdullah, University of Malaysia Terengganu, Malaysia
- Ms. Suneet Kumar, Uttarakhand Technical University/Dehradun Institute of Technology, Dehradun, Uttarakhand, India
- Mr. Rupesh Nasre., Indian Institute of Science, Bangalore., India.
- Mrs. Dimpi Srivastava, Dept of Computer science, Information Technology and Computer Application, MIET, Meerut, India
- Dr. Eva Volna, University of Ostrava, Czech Republic
- Prof. Santosh Balkrishna Patil, S.S.G.M. College of Engineering, Shegaon, India
- Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology Solan (HP), India
- Mr. Ashwani Kumar, Jaypee University of Information Technology Solan(HP), India
- Dr. Abbas Karimi, Faculty of Engineering, I.A.U. Arak Branch, Iran
- Mr. Fahimuddin.Shaik, AITS, Rajampet, India
- Mr. Vahid Majid Nezhad, Islamic Azad University, Iran
- Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore-641014, Tamilnadu, India
- Prof. D. P. Sharma, AMU, Ethiopia
- Dr. Sukumar Senthilkumar, School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia
- Mr. Sanjay Bhargava, Banasthali University, Jaipur, Rajasthan, India

- Prof. Rajesh Deshmukh, Shri Shankaracharya Institute of Professional Management & Technology, India
- Mr. Shervan Fekri Ershad, shiraz international university, Iran
- Dr. Vladimir Urosevic, Ministry of Interior, Republic of Serbia
- Mr. Ajit Singh, MDU Rohtak, India
- Prof. Asha Ambhaikar, Rungta College of Engineering & Technology, Bhilai, India
- Dr. Saurabh Dutta, Dr. B. C. Roy Engineering College, Durgapur, India
- Dr. Mokhled Altarawneh, Mutah University, Jordan
- Mr. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar, India
- Mr S. A. Ahsan rajon, Computer Science and Engineering Discipline, Khulna University, Bangladesh
- Ms. Rezarta Mersini, University of Durres, Albania
- Mrs. Deepika Joshi, Jaipuria Institute of Management Studies, India
- Dr. Niraj Shakhakarmi, Prairie View A&M University, (Texas A&M University System), USA
- Mrs. A. Valarmathi, Anna University, Trichy, India
- Dr. K. Balamurugan, Institute of Road and Transport Technology, India
- Prof. K S Sridharan, Sri Sathya Sai Institue of Higher Learning, India
- Mr. Okumoku-Evroro Oniovosa, Delta State University, Abraka, Nigeria
- Mr. Rajiv Chopra, GTBIT, Delhi, India
- Mr. Harish Garg, Department of Mathematics, IIT Roorkee, India
- Mr. Ganesh Davanam, Sree Vidyanikethan Engineering College, India
- Mr. Bhavesh Shah, VIT, India
- Dr. Suresh Kumar Bhardwaj, Manav Rachna International University, India
- Dr. Muhammad Nawaz Khan, School of electrical engineering & Computer SCience, Pakistan
- Ms. Saranya, Bharathidasan University, India
- Mr. Sumit Joshi, GRD-IMT, Dehradun, India
- Dr. Mohammed M. Abu Shquier, Tabuk University, School of Computers and Information Technology, Kingdom of Saudia Arabia
- Ms. Shalini Ramanathan, PSG College of Technology, India
- Mr. S.Munisankaraiah, Geethanjali college of Engineering & Technology,Hyderabad, India
- Dr. Satyanarayana, KL University, India
- Mr. Sarin CR, Anna University, India
- Mr. Sayed Shoaib Anwar, Mahatma Gandhi Mission College of Engineering, India
- Mrs. Gunjan, JSSATE, Noida, India
- Dr. Ramachandra V Pujeri, Anna University, India
- Mrs. Antima Singh Puniya, Shobhit University, Meerut, India
- Dr. Avdhesh Gupta, College of Engineering Roorkee, India
- Ms. Shiva Prakash, Madan Mohan Malaviya Engg. College, Gorakhpur, India
- Dr. Kristijan Kuk, School of Electrical Engineering and Computer Science Applied Studies, Belgrade, Serbia
- Prof. Dinesh Vitthalrao Rojatkar, Govt. College of Engineering, Chandrapur, India
- Prof. Lalji Prasad, RGTU/TCET, Indore, india
- Dr. A. John Sanjeev Kumar, Thiagarajar College of Engineering, Madurai, Tamilnadu, India
- Mr. Harishbabu Kalidasu, Priyadarshini Institute of Technology and Science, Tenali, Guntur(DT), Andhra Pradesh, India
- Prof. Vaitheeshwaran, Priyadharshini Indira Gandhi College of Engineering, India
- Mrs. P.Salini, Pondicherry Engineering College, India

- Mr. Vivek Bhambri, Desh Bhagat Institute of Management and Computer Sciences,Mandi Gobindgarh(Punjab), India
- Mr. Slavko Zitnik, Faculty of Computer and Information Science Ljubljana, Slovenia
- Ms. Sreenivasa Rao, CMJ University/Yodlee Infotech, India

# TABLE OF CONTENTS

# Efficient Digital Image Authentication and Tamper Localization Technique Using 3Lsb Watermarking

**Sajjad Dadkhah[1], Azizah Abd Manaf [2] and Somayeh Sadeghi[3]**

**[1,3] Faculty of Computer Science and Information System, University Technology Malaysia**
**54100 Kuala Lumpur, Malaysia**

**[2] Advanced Informatics School , University Technology Malaysia**
**54100 Kuala Lumpur, Malaysia**

**Abstract**

The authentications of digital images have become the center of attentions for certain group of companies since the number of doctored images increased. Tampering the digital images in a way that it's impossible to be detected by naked eyes has become easier with development of image editing tools. Digital watermarking can preserve the authentication of digital images. In this paper we proposed an efficient image tamper detection method using 3 LSB (last significant bit) watermarking technique which is able to authenticate the digital image and detect the tamper locations accurately. In the proposed algorithm a 12-bit watermark key will be created from each block of host image which will be embed to last three significant bit of each block. Our proposed method improved tamper detection technique proposed by Prasad's in sense of tamper detection rate by 40 percent. The experimental result clearly proved the efficiency of our proposed method.

***Keywords:*** *Image Authentication, Tamper Detection, 3Lsb watermarking, watermarking for tamper detection.*

## 1. Introduction

With tremendous growths of digital image technology, and low-cost distribution of digital image processing tools among professionals and armature users, preserving the authenticity and integrity of digital images became as a considerable aspect for several organizations [1, 2]. The best way to certify the integrity and authenticity of the digital image is by using digital watermarking. For any kind of watermarking to happen a digital signature will be created to be embedded into particular host image, there are two kind of digital signature visible and invisible. Visible watermarking usually use for the logo of the specific institute or some explanation message of the whole original image which can be a part of the image. Invisible digital signature or watermarked is hidden in original image [3]. The imperceptible watermarking or invisible watermarking has 3 categories which are fragile watermarking, semi-fragile and robust watermark. Fragile watermarking can be broken easily which means it's not totally immune from blind attacks. Fragile watermarking methods are very weak in functional sense. However because fragile watermarking is very sensitive to any kind of manipulation it can be very valuable for authentication purposes [4, 5]. Moreover Semi-fragile watermarking techniques are very suitable for tamper localization. This method of watermarking can be used to locate any alternation and extra modification that occurred inside an image [6]. Authentication or tamper detection for digital image is to perceive any sort of manipulation inside the particular image and distinguish the original image from the fake one [7, 8]. The digital watermarking can be divided into block-wise and pixel-wise techniques. Block-wise watermarking is dividing the host image into specified non overlapping blocks for the purpose of tamper detection [9].

Furthermore researchers started to pay more attention to digital watermarking area. Fridrich [10] proposed a fragile watermarking method which could protect against Vector Quantization attack (VQ attack). Quantization attack target independent block-wise approaches by using the host image information. Wong [11] proposed a public-key fragile watermark which divides the image into non-overlapping blocks, then embeds a digital signature into last significant bit of each block. Chang et al. [12] conducted series of attacks to examine a hierarchical digital watermarking method for image tamper detection proposed by Lin et al. [13]. The experimental result of

their scan illustrated that parity check and intensity-relation of the image blocks cannot help to detect all types of tampering attacks.

However Lee and Shinfeng [14] proposed a Dual watermarking method for image tamper detection and recovery that solved some basic problem like no second chance for watermarked information, in case watermarked information destroyed by tampering in. One year later Chaluvadi and Prasad [15] highlighted very important flaw in Lee and Shinfeng's method. They proved that lee and Shinfengs method cannot detect any tampering attacks in 5 MSB (most significant bit) of watermarked image.

To evaluate tamper detection method, various attacks can be performed. In order to achieve robustness, particular tamper detection methods have to be capable to detect collusion attacks [16]. Collusion attack can be deleting some portion of the watermark image or copy some part of the watermark image and paste it anywhere inside the same image. Performing bit tampering attack which is modifying a specific bit in specific position, can illustrate the consistency of any tamper detection. Tamper detection rate can be calculated in order to evaluate the capability of proposed authentication method. The tamper detection rate in our case is the number of tampered blocks detected by proposed method which will be explained in following sections.

The remaining of paper is organized as follows. An overall analysis of Chaluvadi and Prasad's dual tamper detection will be conducted in section 2. In section 3 the methodology of our proposed image authentication using 3 LSB) will be explained. Section 4 illustrated the examination result of the proposed method along with the comparison of result with two current methods. Finally we conclude our proposed method in section 5 and explain the contribution of proposed method.

## 2. Prasad's Tamper Detection Method

They used 3-level hierarchical tamper detection proposed by Lin [17] which is block-based tamper detection method. Their system has both tamper detection and recovery capability. The system starts with grayscale image with MxM size, where M assumed to be multiple of 2. First they divide the whole image to non-overlapping blocks of size 2x2. Their approaches consist of two parts, first embedding the watermark and second the 3-level tamper detection technique.

The watermarking approach starts with constructing a 12-bit watermark that is going to be embedded in last three significant bit of each block. Their method divides the image horizontally into two equal parts and select two blocks from same position. The 12-bit watermark is consisting of representative block information and average intensity of each block which is simply the average of 4 pixels inside each block. The last 2-bit of their watermark is the parity check of the average intensity which is number of one's inside the binary form of particular bit, if the number of ones were even the parity check value will be 0 and if the number of ones were odd the parity check will be 1.

$$r = \begin{cases} 0 & \text{if count is even} \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

Their proposed method included a smoothing function which improve the quality of the watermark and recovered image. To evaluate their tamper detection, they have conducted two bit tampering attacks. We conducted exactly the same attacks as illustrated in [15] to be able to compare results. The experiment results are illustrated in section 4.

## 3. Proposed approach

The proposed algorithm of the system includes 3LSB watermarking which is same as [1] and proposed 2-level tamper detection method. The overall procedure of the proposed system is illustrated in Fig. 2.

However as figure 1 illustrated after dividing the image into blocks of 2x2 and generating the 12-bit watermark, the 12-bit watermark will be embedded into 3LSB of each block. The proposed tamper detection method is consist of two parts which both are the comparison of the content of the 12-bit that has been watermarked to host image .In the first level the average intensity bits will be compared which is the last 2 bits of the 12-bit, second level is the comparison of the remaining 10 bit which all has to be identical otherwise tampered has happened. The 2-Level of our tamper detection method will make sure of the efficiency and accuracy of tamper detection and localization. The procedure of proposed 3LSB watermarking and 2-Level tamper detection will be explained in section A and B.

### 2.1 Watermarking Procedure

In this phase, we identified and generated a 12-bit watermark for embedding into last three significant bit of each block. The watermark content which has 12-bit size have to be something that relate to each pixel of the block so in case of tampering in one of the pixels, our method can detect and localize the tampered blocks. The first step in watermarking procedure is constructing the 12-bit watermark. The procedure of constructing the 12-bit watermark for our proposed system is same as [15] in primary steps but the contents of the watermark is different, which improved the tamper detection rate as illustrated in experiment results.

First we pad the 3 last significant bit of each pixel to zero. It will produce a new value obviously, thus we calculate the average intensity of the new value which is simply the average of the 4 pixels inside the block followed by calculating the parity check with (1).

So far, one bit of 12-bits has been generated and 11 bits remain. The main purpose of generating these 12-bits watermark is to make a digital signature for each block of size 2x2. Thus for generating the 12-bit watermark as figure 2 illustrated the 5MSB (most significant bit) of watermark is 5MSBof the first pixel inside the block and the second 5MSB is selected from the second pixel of the block.

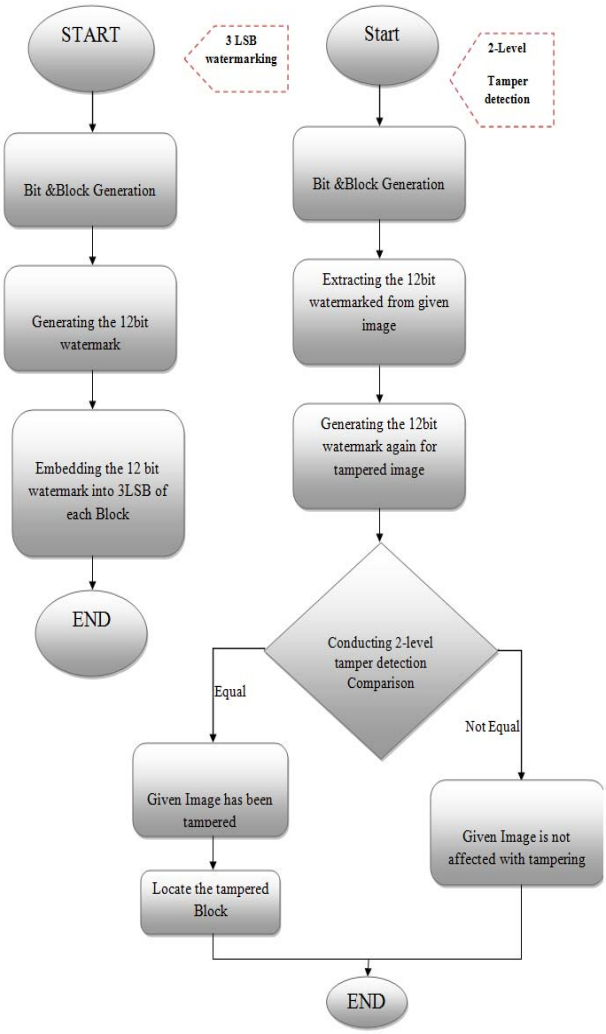



Fig. 1 General Structure of Proposed algorithm

After generating the 11 bits of watermark the best constant value that can be useful for tamper detection is parity check of the 10-bit that we just constructed in the previous section. As illustrated in figure 2 the parity check of the 10 bit calculated by (1) name as Cr. The last bit of the watermark is r which is the parity check of average intensity of the whole block.

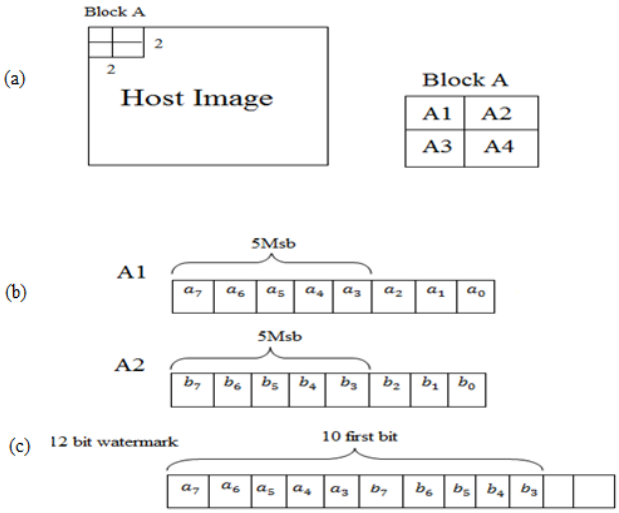



Fig. 2 Generation 10 MSB of watermark :(a)the block of size 2x2, (b) 5MSB of the first and second pixel of the block, (c) 10 bit of 12-bit watermark

As illustrated in figure 3 the first 5 bit of 12-bit of watermark will be embed to first two pixels inside each block, and the next 5 bit of watermark will be embedded into pixel three and pixel four. The last 2 significant bit of the watermark pixel will be embedded to 2nd LSB of the pixel number four inside each block. This embedding will be done for entire blocks inside the host image.

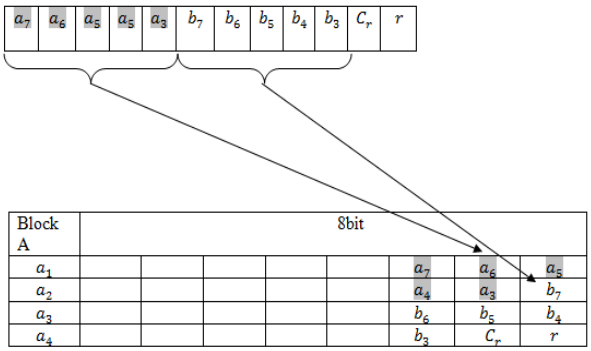



Fig. 3 Proposed watermarking process

### 2.2 Proposed Authentication Procedure

As figure 1 illustrated, the main concept of proposed tamper detection method is two important comparisons. So two level of tamper detection will be perform. If the tampered is not detected in fist level of tamper detection, the second level makes sure to detect the tampered blocks. After dividing the watermarked image to blocks of size 2x2, the last three significant bit of each pixel within the image will be padded to zero. The average intensity and parity check of new pixel value will be calculated and will be compared to last significant bit of the extracted watermark. This comparison will be done for each block of size 2x2 separately if the both value were equal then there has not been any tamper occurrence inside the image otherwise the tamper has happened and the method will mark that block as invalid to localize the tampered area. The first level of our tamper detection technique continues with calculating the parity check of the first 10 bits of the extracted watermark and compares the value to the second LSB of the same extracted watermark, If both values were equal to each other no tamper has happened, otherwise the block that method is doing comparison concurrently will be marked as invalid for purpose of tamper localization.
The efficiency of proposed tamper detection method will be guarantee with second level of tamper detection technique. After the first level if the method detects any tamper it would mark the tampered block as invalid and go on to second level. In this level a bit by bit comparison will be conduct between the 5MSB of the first two pixels and first 10 bit of the extracted watermark as illustrated in figure 1.

## 3. Experimental result

The first tampering attack that will be performed on proposed system is one type of collage attack. The collage attack that is going to be performed is made by copying one region from the watermarked image and pasting the region somewhere inside the same watermarked image. The second type of tampering attack is deletion attack which is simply deleting some part of watermarked image, however this attack will be performed in different scales to evaluate the proposed system ability in detecting the tampered areas.
The next tampering attack that is going to be performed is tampering with watermarked pixels bits. In order to compare the result of proposed system with recent tamper detection methods, attacks that are going to be perform in this phase of system, have to be same attacks as conducted in Chaluvadi and Prasad's tamper detection method [15]. The gray scale Lena image with size of 256x256 will be used for this experiment. According to size of testing image, total number of blocks with size 2x2 will be 16384. For conducting the same experiment as Chaluvadi and Prasad's method 10 % of total blocks with difference of 10 blocks between each selected block have to be selected.
The first type of bit tampering attack is Attack-1 which is made by changing the 4th LSB of selected blocks to value 1. It means if the value of the 4th LSB is 0, will be changed to 1 and if the value of the 4th LSB is already 1, no change will be performed. So in this attack number of selected blocks for tampering and number of actual tampered blocks will be different.
The second type of bit tampering attack that has been performed is Attack-2 which is made by modifying the 4th LSB of pixels inside the selected blocks by replacing 0 with 1 and 1 with 0 for all the 4th LSB inside the blocks. Thus the number of selected blocks and actual number of tampered blocks in this attack will be equal to each other, which make this attack a perfect experiment for evaluating the proposed tamper detection system. However tamper detection rate will be calculated by (2).

$$Dt = \frac{number\ of\ detected\ blocks}{Number\ of\ tampered} X\ 100 \qquad (2)$$

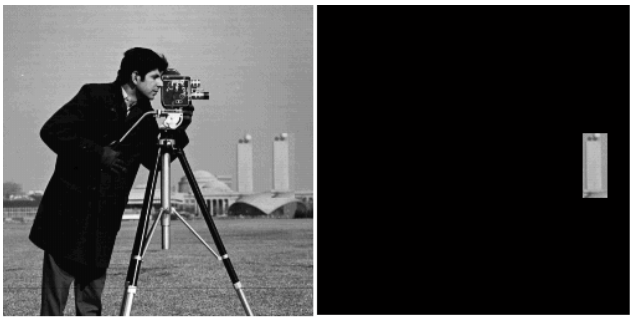
### 3.1 Collage Attack experimental result

The following figures illustrating the collage attack on gray scale images. The attack is copy and pastes some part of the image inside the image. As illustrated in following Figures the proposed method accurately detected the tampered areas.

(a) (b)

Fig. 4 Collage Attack: (a) tampered room, (b) Tamper detected

To achieve a good set of experiment, deletion attack is going to be performed in small and large size with different set of images. Fig. 9, Fig. 10 and Fig. 11 clearly illustrated the power of our proposed tamper detection method in detecting collage tampers in any size and scale. These images illustrate deletions in different size which all have been detected by system tamper detection method. Moreover, the localization of tampered blocks also has been calculated by proposed system.



(a) (b)

Fig. 5 Collage Attack2: (a) tampered camera man, (b) Tamper detected

(a) (b)

Fig. 6 Collage Attack3: (a) tampered house, (b) Tamper detected

As figure 5 and figure 6 illustrated the proposed authentication method is capable to detect the copy-move collage attack accurately.

## 3.2 Experiments Results of Deletion Attack

To achieve a good set of experiment, deletion attack is going to be performed in small and large size with different set of images. Figure 7, figure 8 and figure 9 clearly illustrated the power of our proposed tamper detection method in detecting collage tampers in any size and scale. These images illustrate deletions in different size which all have been detected by system tamper detection method. Moreover, the localization of tampered blocks also has been calculated by proposed system.

## 3.3 Experimenting result of Attack-1and Attack-2

In this section tampering on the 4th LSB of 10 % of selected block will be performed, the result will be used in comparison with [15] and [14]. The selected blocks are chosen from different position within the host image. Figure 10 illustrates performing of attack-1 on gray scale Lena image with size of 256x256. As figure 10 illustrated, proposed tamper detection method were able to detect 1187 of 1458 modified blocks. The total number of tampered blocks in this attack supposes to be 1639 blocks, but because Attack-1 attempted to replace all 4th LSB of watermarked image with 1 the bits that already have value 1 won't be changed.

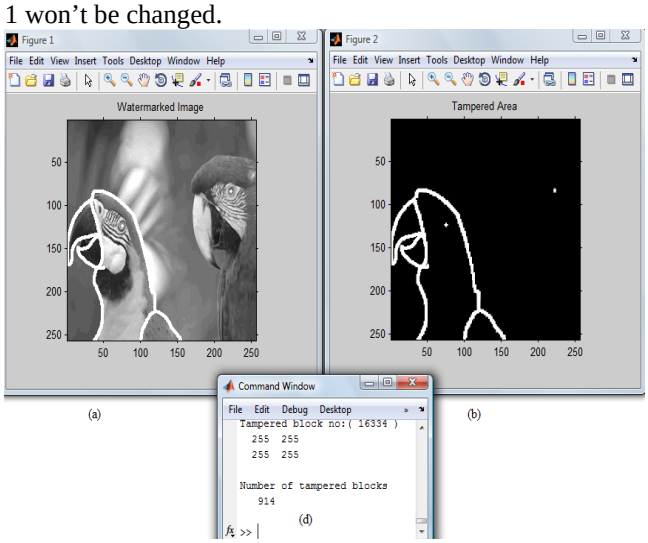



Fig. 7 DeletionAttack1: (a) tampered parrot, (b) Tamper detected, (d) number of tampered blocks

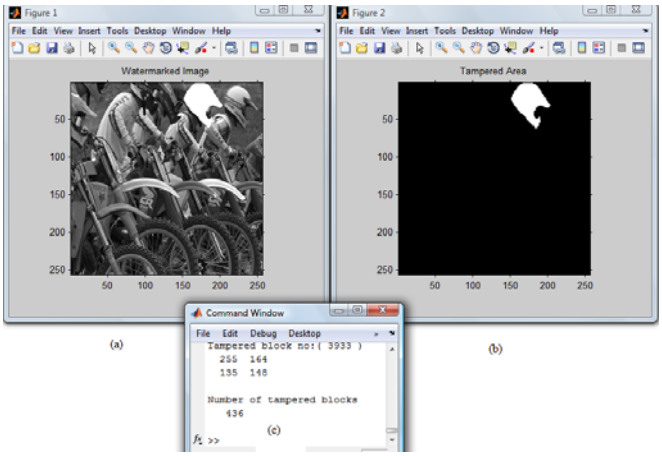



Fig. 8, DeletionAttack2: (a) tampered bike, (b) Tamper detected, (d) number of tampered blocks

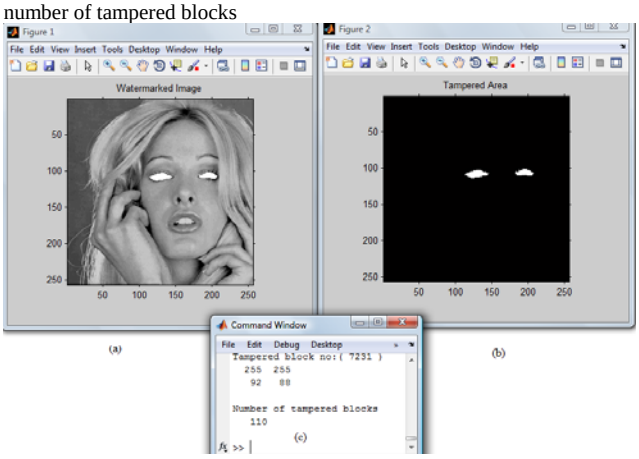



Fig. 9, DeletionAttack3: (a) tampered face, (b) Tamper detected, (d) number of tampered blocks

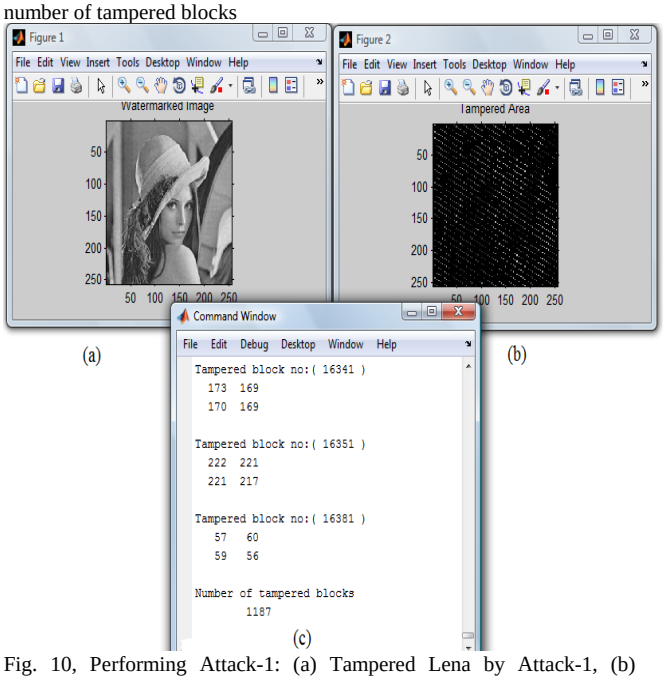



Fig. 10, Performing Attack-1: (a) Tampered Lena by Attack-1, (b) Tampered area detected, (c) Number of tampered blocks detected

Attack-2 has been performed on 10% of selected blocks. As explained previously, this tampering attack modifies the 4th LSB of the watermarked image. So if the value of the 4th LSB were 1, attack will change it to 0 and if the value were already 0 it will be changed to value 1. Attack-2 is perfect experiment to evaluate tamper detection rate, because number of modified blocks is equal to the number of actual tampered blocks which in this case is 1639 blocks.

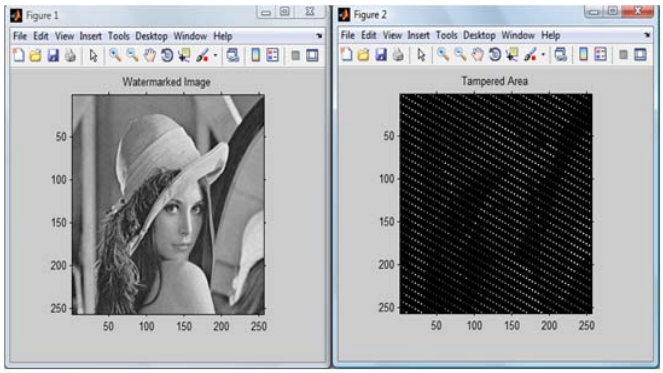



Fig. 11, Performing Attack-2: (a) Tampered Lena by Attack-2, (b) Tampered area detected

Figure 11 illustrates performing attack-2 on gray scale Lena image with 256x256 pixel size. As the figures illustrated our proposed tamper detection method was able to detect all the tampered blocks successfully which is 1639 blocks. Moreover, the system presents the tampered areas graphically and localizes all the tampered blocks with the corresponding blocks numbers.

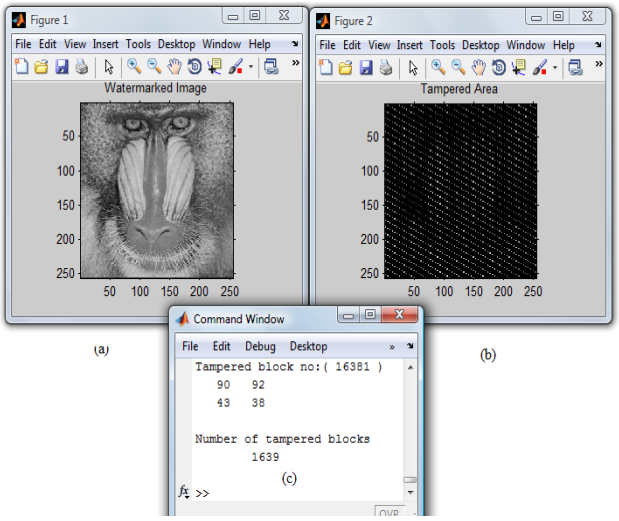



Fig. 12, Performing Attack-2 on Baboon: (a) Tampered Baboon by Attack-2, (b) Tampered area detected

Figure 12 illustrates performing attack-2 on watermarked Baboon image. As it is shown in figure 12 all the tampered blocks has been detected by the proposed authentication system. The number of tampered blocks is 1639 block which is exactly the same as number of detected blocks by the proposed system.

## 4. Evaluating the proposed tamper detection

A dual tamper detection method proposed by Lee and Shinfeng's [14] have been improved by Chaluvadi and Parsad's [15] tamper detection method and our proposed 2-Level tamper detection has improved the tamper detection percentage of [15]. The comparison of Attack-1 and Attack-2 experiment result is illustrated in Table 1.

Table 1: Comparison of tamper detection percentage

| # | ***Attack-1*** | ***Attack-2*** |
|---|---|---|
| 10% of Blocks selected in watermarking | **1639** | **1639** |
| Modified blocks in Watermark Image (%) | 1458(1458/ 1639 = **88.96%)** | 1639 (**100%)** |
| No of Detected blocks by Lee and Shinfeng's method [14] (%) | NIL (**0%**) | NIL (**0%**) |
| No of Detected blocks by Chaluvadi and Parsad's method [15] (%) | 649 (649/1458 = **47.59%**) | 849 (849/1639= **51.79%**) |
| No of Detected blocks by system's proposed method | 1187(1187/1458 = **81.41%)** | 1639 (1639/1639= **100 %** ) |

After conducting the Attack-1 on Lena watermarked image, Lee and Shinfeng's method were not able to detect even one single tampered block, Chaluvadi and Parsad's method [15] were able to detect 649 blocks which is 47.59 % of tampered blocks, but our proposed 2-Level tamper detection method detected 1187 of tampered blocks which is 81.41 % of tamper detection percentage which prove the efficiency of propos system. The number of selected blocks for Attack-2 is 1639 which is exactly equal to number of actual modified blocks. Lee and Shinfeng's method still were not able to detect one single block for this specific attack and Chaluvadi and Parsad's method [15] were able to detect 849 tampered blocks which is 51.79 % tamper detection rate. Finally our proposed tamper detection method had 100% detection rate for this high level attack which proved the high efficiency and accuracy of our proposed method.

## 5. Conclusions

In this paper, we illustrated an efficient method for image authentication and tamper localization. Different types of tampering attacks have been experimented in order to evaluate the proposed method. The examination results of our proposed authentication obtained high tamper detection rate result. However the evaluation result of proposed system proved the efficiency and accuracy of the proposed authentication method in detecting and locating the tampered area.

### Acknowledgments

The authors would like to thank University Teknologi Malaysia for their educational and financial support.

## References

[1] W. Lin et al, "Multimedia Analysis, Processing & Communications", Springer-Verlag Berlin Heidelberg, 2011 SCI 346, pp. 139–183.
[2] G. Friedman, "The trustworthy digital camera: restoring credibility to the photographic image ", Proceedings of the IEEE International Conference on Image Processing. Chicago, IL, USA. October 1998. pp. 409–413.
[3] Y. Tien, and D.Shinfeng, "Dual watermark for image tamper detection and recovery, Pattern Recognition", 2008. pp.3497–3506.
[4] D. Kundur,. and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication", Proceedings of the IEEE, 1999, Vol. 87, pp. 1167–1180.
[5] G.Yu, and H. Liao, "Mean quantization-based fragile watermarking for image authentication", 2001, pp. 1396–1408.

[6] C.Y Lin, .and C. Chang, "Semi-fragile watermarking for authenticating JPEG visual content", SPIE International Conference on Security and Watermarking of Multimedia Contents II.San Jose, USA, January 2000.
[7] S.H. Liu, H.X. Yao, W. Gao, and Y.L. Liu, "An image fragile watermark scheme based on chaotic image pattern and pixel-pairs", Applied Mathematics and Computation, 2007, vol. 185, no. 2, pp. 869-882.
[8] C. Vleeschouwer, J.F. Delaigle, and B. Macq, "Invisibility and application functionalities in perceptual watermarking-an overview", Proceeding of the IEEE, 2002, vol. 90, pp. 64–77.
[9] H. H, N. Zhang, and H.M. Tai, "Awavelet-based fragilewatermarking scheme for secure image authentication", Proceeding of 5th International Workshop on Digital Watermarking.
[10] J. Fridrich, "Security of fragile authentication watermarks with localization", Proceeding of SPIE Security and Watermarking of Multimedia Contents IV, San Jose, CA, 2002, vol. 4675, pp. 691–700.
[11] P.W. Wong, "A public key watermark for image verification and authentication", Proceedings of the IEEE International Conference on Image Processing, 1998, vol.1, pp.455-459.
[12] C. Chang, Y. Fan, and W. Tai, "Four scanning attack on hierarchical digital watermarking method for image tamper detection and recovery", Pattern Recognition, 2008, vol. 41, pp. 654-661.
[13] S.D. Lin, Y. Kuo, and M, Yao, "An image watermarking scheme with tamper detection and recovery", International Journal of Innovative Computing, Information and Control, 2007, vol. 3, no. 6(A), pp.1379-1387.
[14] T. Lee and S. D. Lin, "Dual watermark for image tamper detection and recovery,"Pattern Recognition", 2008, vol. 41, pp.3497–3506.
[15] S.B. Chaluvadi, M.V.N. Prasad, "Efficient Image Tamper Detection and Recovery Technique using Dual Watermark", Proc. IEEE. World Congress on Nature & Biologically Inspired Computing (NaBIC09), IEEE Press, Dec. 2009, pp.993, doi:10.1109/NABIC.2009.5393888.
[16] D. Kirovski, H.S. Malvar, and Y. Yacobi, "A dual watermarking and fingerprinting system", IEEE Multimedia, 2004, vol. 11, no. 3, pp. 59–73.
[17] P.L. Lin, C.K. Hsieh, and P.W. Huang, "A hierarchical digital watermarking method for image tamper detection and recovery", Pattern Recognition, 2005, vol. 38, pp. 2519–2529.

# A Method using Language Grid and Concept Base for Japanese-English Cross-language Information Retrieval

**Pham Huy Anh[1]and Yukawa Takashi[2]**

**[1] Department of Information Science and Technology, Nagaoka University of Technology, Nagaoka-shi, 940-2188 Japan**

**[2] Department of Information Science and Technology, Nagaoka University of Technology, Nagaoka-shi, 940-2188 Japan**

**Abstract**

This paper describes query translation using language resources and a concept base method for Cross-language Information Retrieval (CLIR). In the proposed method, queries are translated by multiple machine translation systems on the Language Grid. The queries are then expanded by using a bilingual dictionary to translate compound words or word phrases. In addition, documents related to the translated query are retrieved with a TF-IDF term weighting model. The top 100 retrieved documents are re-ranked by a specificity-considered concept base with the noun phrases and compound words extracted from the query. The re-ranked results are combined with the results retrieved by the probabilistic model. For evaluation of the proposed method, we use the average precision of the non-interpolated recall and precision to compare our method with the NTCIR1 participation systems. The proposed method achieved the highest precision.

***Keywords:*** *Cross-language Information Retrieval, CLIR, Language resources, Concept base, Language Grid.*

## 1. Introduction

The number of electronic documents on the Internet has rapidly increased. As a result, documents containing the kinds of information required by a user are not limited to those written in the user's native language. Therefore, research on Cross-language Information Retrieval (CLIR), which uses a query in one language to retrieve documents in another language, is especially of interest. In the NTCIR (NII-NACSIS Test Collection for IR Systems) workshop, [1] CLIR is one task that is being specifically investigated by various organizations. To retrieve information in other languages, a query is translated by the machine translation system. Therefore, not only a retrieval model but also language resources and language processing functions are important factors of the CLIR system. Information retrieval models include the probabilistic model proposed by Robertson and Sparck Jones [2] and the vector space model proposed by Salton and McGill [3], both of which are used in CLIR.

Language resources include a dictionary, thesaurus, and bilingual corpus. Language processing functions include, for example, morphological analysis and machine translations.

To increase the performance of CLIR, a highly accurate query translation system is constructed by using existing multiple translation systems and improving the retrieval model. Limitations in the dictionary vocabulary and the presence of word ambiguity pose problems for query translation. Although numerous research studies have used the sentence translation system and the bilingual dictionary for query translation, the problems of such query translation remain unsolved.

We propose query translation using multiple machine translation systems on the Language Grid. Two machine translation systems are utilized in this method for translating a query, and a bilingual dictionary is used for translating the compound words and noun phrases of a query. To overcome the problem of mistranslated words appearing in the query, a filtering method using the concept base is proposed. In particular, to delete a mistranslated word, the similarity between the back translation of the word and the word in the source query is calculated. A word having a low similarity is considered to be the mistranslated word. In this paper, we describe "Using Language Grid for CLIR" method, "Deleting mistranslated queries using improved concept base" method and "Re-ranking retrieve results using concept base" method in detail, discuss its implementation, and present its experimental evaluation.

## 2. Related work

Cross-language information retrieval is divided into the query translation part and the information retrieval part. In the query translation part, many research studies have focused on a method using a sentence translation system and a bilingual dictionary [4] [5].

Atsushi Fujii and Tetsuya Ishikawa proposed a method integrating query and document translation using Machine Translation (MT) [6]. Aitao Chen et al. proposed a method combining multiple sources for short query translation in Chinese-English using two transfer dictionaries [7]. Wang et al. proposed a method of dictionary expansion using Wikipedia [8]. However, all of these methods had limitations. If only the bilingual dictionary was used for query translation. As a result, the ambiguity problem remained unsolved. In addition, if a query was expanded after translation, a mistranslated word could also be expanded. In addition, MT translates a query based on the context of the query. In MT, an input sentence is translated into an output sentence. If MT mistranslates the query, the mistranslated word appears in the query. The query then produces an inaccurate retrieval result.

We propose query translation using multiple machine translation systems and a bilingual dictionary on the Language Grid. In addition, to delete the mistranslated word in the query, the concept base is used. All language resources used in the method can be utilized in the Internet for CLIR system.

## 3. Background

To perfect the performance of CLIR, the accuracy of the query translation and the information retrieval must be improved.

The Language Grid is a new multilingual infrastructure on the Internet available for intercultural collaboration. This system can be used by freely combining the language resource and the language processing function in the Internet. By combining the multiple language resources on the Language Grid, it is possible to produce a translation result having high accuracy.

In the proposed method, the multiple machine translation systems and a bilingual dictionary open to the public in the Language Grid are combined, and highly accurate cross-language information retrieval is achieved. However, a mistranslated word in a query increases by using multiple language resources. Accordingly, the filtering method using a concept base is proposed for deleting the mistranslated word.

### 3.1 Language Grid

To satisfy the needs of users, the Language Grid allows users to easily develop new language services by combining existing ones.

The development of Semantic Web technologies enables the collaboration needed among language resources and language processing functions. The language resources include bilingual dictionaries, thesauruses and corpora, and the language processing functions include machine translation, morphological analysis and paraphrases. The Simple Object Access Protocol (SOAP) has been used for accessing the language resources of the Language Grid. SOAP specifies the exchange of structured information in the implementation of web services in computer networks. Web Services Description Language (WSDL) is a specification that describes networked XML-based services. WSDL provides a simple way for service providers to describe the basic format of requests to their systems regardless of the underlying SOAP protocol.

The Language Grid service layer includes the peer-to-peer (P2P) grid infrastructure, the language resource, the language service, and the intercultural collaboration tools. In our research, the retrieval service for CLIR is constructed in the layer of the P2P grid infrastructure and the language resource. Although multiple languages can be translated by the Language Grid, our data set is Japanese and English, so only a Japanese-English resource is needed.

### 3.2 Concept base

To delete the mistranslated word in a query, the concept base is used. In addition, concept base re-ranking is conducted with a noun phrase and a compound word extracted from the query.

The concept base was proposed by Schüetze and Pederson as a method of automatically constructing a thesaurus with the corpus and using a higher dimension vector space to express the relations between words appearing in a document [11]. Currently, the commonly utilized composition of the concept base is a word × word matrix.

First, in the traditional construction of the concept base, N words occurring with high frequency in the document for retrieval are selected to create a neighborhood co-occurrence matrix of a word with another word in the neighborhood (Figure1). Wij is the co-occurrence frequency between word i and word j. Before constructing the neighborhood co-occurrence matrix, it is necessary to perform a morphological analysis and remove the stopwords as a preprocessing step. Stopwords are words such as particles or auxiliary verbs that does not have an important role in the documents.

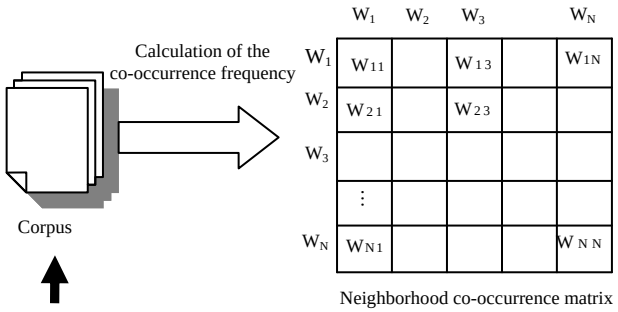



Fig. 1 Making the neighborhood co-occurrence matrix.

The created neighborhood co-occurrence matrix can be considered as a word vector in which the number of words corresponds to the number of dimensions. However, there is the problem that the number of dimensions increases as the scale of the corpus grows because dimension depends on the number of words. Moreover, because each axis is a word, it is not easy to think of the axes as being mutually orthogonal. Therefore, to create the neighborhood co-occurrence matrix, singular value decomposition (SVD) is implemented. Under SVD, the neighborhood co-occurrence matrix is divided into three matrices: the transposed orthogonal matrix, the diagonal matrix, and the row orthogonal matrix. The row of 100~200 dimensions is extracted from the obtained row orthogonal matrix. The extracted matrix is the concept base (Figure 2).

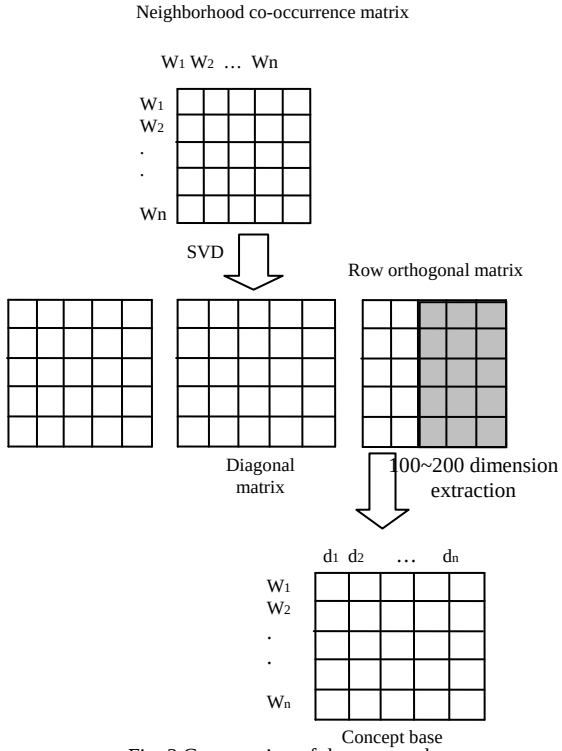



Fig. 2 Construction of the concept base.

## 4. CLIR System using the Language Grid and the Concept Base

### 4.1 System overview

Figure 3 depicts the overall design of our CLIR system. First, "Tokenizer E" processes the English language documents ("Doc in E") and "Tokenizer J" processes the Japanese language documents ("Doc in J"). If the documents are in English, the TreeTagger morphological analyzer and discard stopwords are utilized. In contrast, Japanese documents are segmented into lexical units using the ChaSen morphological analyzer and discard stopwords. Thereafter, the concept base is made from outputs of the tokenizer.

In the "Translator," a source language query ("Query in E") outputs the translation ("Query in J") by using the language resources of the Language Grid. We used two "Machine translations" to translate a query sentence and a "Dictionary" to translate the compound words and noun phrases of the query. Then, in the "Mistranslated word deletor," the "Concept base" is utilized for deleting the mistranslated words.

Finally, the "IR engine" outputs the top 1000 documents in descending order according to the similarity between the translated query and each document. The "Result combiner" combines the output of the two different IR engines. The first engine ("TF-IDF model") is a naive implementation of the vector space model with Term Frequency-Inverse Document Frequency (TF-IDF) term weighting, from which "Result re-rank" uses the concept base including the noun phrases for re-ranking the outputs of the top 100 documents. The second engine ("Probabilistic model") is a probabilistic model of the Lemur retrieval system.

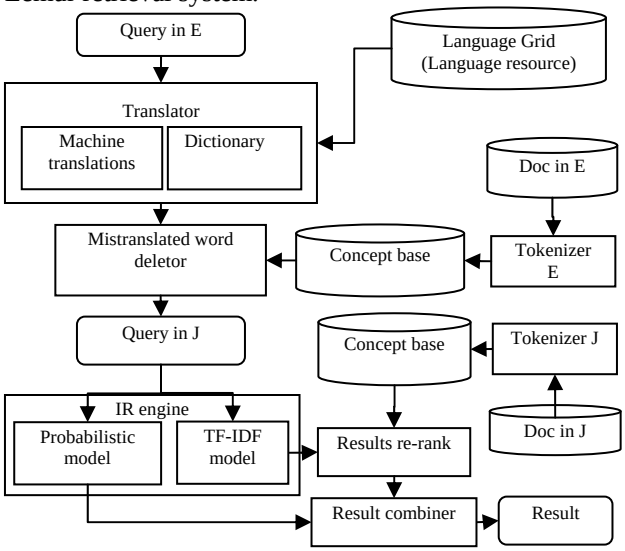



Fig. 3 The overall design of our CLIR system.

### 4.2 Query translation and expansion using language resources

In query translation, three unresolved matters remain: mistranslated words, limitations in the dictionary vocabulary and the ambiguity problem. This paper proposes query expansion using multiple machine translation systems on the Language Grid with a bilingual dictionary to compensate for the limited vocabulary and ambiguity problems. Specifically, for translation of a given query in the source language, we use the two

machine translation systems in the Language Grid, one of which reduces the ambiguity problem. Moreover, the query is expanded to compensate for the vocabulary limitations of the dictionary. In a query, compound words and noun phrases play important roles in deciding the retrieval result. Therefore, compound words and noun phrases are used for query expansion, and they are translated by the bilingual dictionary. We extract compound words and noun phrases from the query in the following example:

"マルチキャスト通信における関連する複数データの品質制御手法について論じたものはないか。"

For this query, we extract only "マルチキャスト通信", "複数データ", and "品質制御手法".

To overcome the problem of mistranslated words appearing in the query, a filtering method using the concept base is proposed.

## 4.3 Deletion of mistranslated query words

By using the language resources for query expansion, mistranslated words can appear within the query. Therefore, after completing the back translation, mistranslated words are found through the similarity of words. It is thought that if the mistranslated words are deleted, a better query can be obtained. Accordingly, a filtering method using a concept base is proposed.

In particular, to delete a mistranslated word, the similarity in the back translation word and the word in the source query is calculated (Figure 4). The source query vector Wi and the back translation word vector Wj are obtained from the concept base. The scalar product in the two word vectors is a degree of similarity of the words.

A word having a low degree (0.8) of similarity is considered to be the mistranslated word, which is then deleted. To delete the mistranslated word, the concept base is improved, as follows.

The concept base is made by using a corpus. The corpus is segmented by the morphological analyzer and stopwords, nouns, noun phrases, and adjectives are extracted.

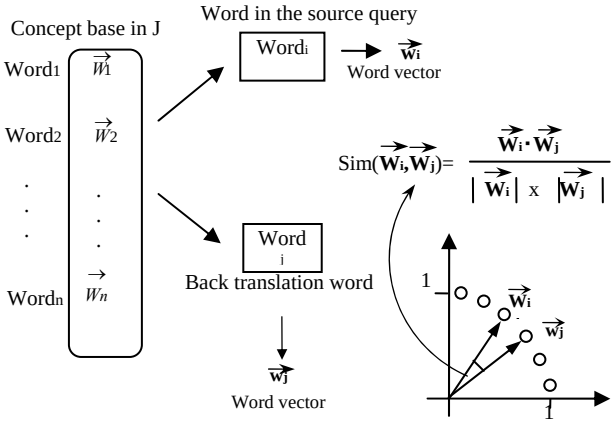



Fig. 4 Calculating method of similarity between a query and a back translated query

Since the concept base is composed of the co-occurrence frequencies between words, it does not consider the specificity element of a document containing a word. Consequently, the retrieval performance is decreased. We propose to construct a concept base in which this specificity is considered. The IDF is an index showing the specificity of a word. The concept base considers the frequency of a pair of words related to the co-occurrence as a single element. The IDF of a word pair is evaluated and used as a weighting term. For example, since the co-occurrence frequency of (computer, network) is greater than one, the IDF of the word pair is calculated as follows (eq. (1)):

$$idf\{pair(t_1,t_2)\} = \log \frac{N}{df\{pair(t_1,t_2)\}} \quad (1)$$

Here, ***pair*** $(t_1,t_2)$ is a pair of the words $t_1$ and $t_2$ that exist in the co-occurrence relation, **N** is the total number of documents, and $df\{pair(t_1,t_2)\}$ represents the number of documents in which the word pair appears.

The $idf\{pair(t_1,t_2)\}$ value becomes the origin of the concept base. The weight is calculated by multiplying this value by each element of the neighborhood co-occurrence matrix. Therefore, element ***W*** of the neighborhood co-occurrence matrix is determined as follows (eq. (2)).

$$W_{t_1t_2} = F_{t_1t_2} \times idf\{pair(t_1,t_2)\} \quad (2)$$

Here, $F_{t1t2}$ is the co-occurrence frequency of words $t_1$ and $t_2$. The element of the neighborhood co-occurrence matrix in Fig. 1 replaces **Wi (i=1,…,n)** with the value of the above expression. The concept base, composed of the neighborhood co-occurrence matrix with the element of eq. (2), is the specific concept base.

The evaluation of how many mistranslated words deleted by the filtering using the concept base was performed. The words matching with the requirement of the retrieval is subjectively evaluated by human. Human examines the word in the query one by one. The fact that whether the word is appropriate to information retrieval is judged. The evaluation results are shown in the Table 1. Before filtering, the average number of mistranslated word of all queries is 6. The average number of mistranslated word of all queries after filtering is 0.417. Numbers of mistranslated word were deleted. The mistranslated word rate deleted by the concept base is 93.05%.

Table 1: Mistranslated word rate is deleted by concept base

| | |
|---|---|
| Average number of mistranslated word of all queries before filtering | 6.0 |
| Average number of mistranslated word of all queries after filtering | 0.417 |
| **Mistranslated word rate is deleted by concept base** | **93.05%** |

### 4.4 Re-ranking using a concept base

By using multiple machine translation systems, a query can contain many words that are the keywords of relevant documents. However, multiple systems also increase the number of words that are keywords of non-relevant documents, and, consequently, the retrieval performance decreases. We carry out re-ranking using a concept base with the noun phrases and compound words extracted from the query. Generally, when information retrieval is performed by the concept base, calculation of the similarity of the query vector and the document vector is carried out. However, if the number of words in the document is much larger than the number of words in the query, the similarity influence and the retrieval performance decrease. In our method, the document vector ($\mathbf{D}$) is divided into sentence vectors ($\mathbf{S}i$). The number of words in the query vector ($\mathbf{Q}$) and the number of words in the sentence vectors are almost the same. The similarity of the query vector and the sentence vector is then calculated. The highest similarity is assumed to be the similarity of the query vector and the document vector. The formula is shown below (eq. (3), (4)).

$$\vec{\mathbf{D}} = \vec{\mathbf{S}}_1 + \vec{\mathbf{S}}_2 + \ldots + \vec{\mathbf{S}}_n \quad (3)$$

$$\mathbf{Sim}(\vec{\mathbf{Q}}, \vec{\mathbf{D}}) = \max_{i=[1:n]} (\mathbf{Q}, \vec{\mathbf{S}}i) \quad (4)$$

where n is the number of sentences in the document.

In addition, the re-rank system is evaluated in comparison with the system which is not re-ranked. Table 2 shows the evaluation results. The non-interpolated average precision values of averaged over 39 queries are compared with 50, 100, 200, 400, 600, 800, 1000 of retrieved documents. All the comparison results of the re-ranking system are exceeded. Re-ranking using a concept base method was effective.

Table 2: Comparison of the results of the re-rank system and the result combined with no re-rank (Non-interpolated average precision values, averaged over the 39 queries)

| Method | | Result combined with no re-rank | Re-rank |
|---|---|---|---|
| Number of retrieved documents(N) | 50 | 0.1740 | 0.1954 |
| | 100 | 0.1829 | 0.2058 |
| | 200 | 0.1876 | 0.2106 |
| | 400 | 0.1896 | 0.2128 |
| | 600 | 0.1905 | 0.2137 |
| | 800 | 0.1907 | 0.2140 |
| | 1000 | 0.1909 | 0.2142 |

## 5. Performance Evaluation

### 5.1 Baseline system

The proposed system is evaluated in comparison with the baseline system, which corresponds to the single machine translation system shown in Figure 5.

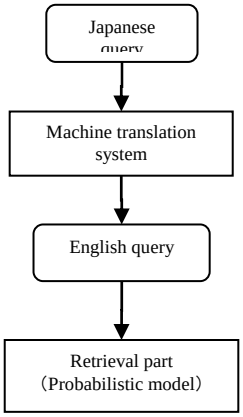



Fig. 5 The baseline system.

## 5.2 Experimental data and environment

We used the test collection of Japanese-English science documents used in the NTCIR1 CLIR task (Japanese documents: 330,000, English documents: 190,000, Japanese queries: 39). To retrieve information, an English document set, from which only a verb and a noun were used for calculating the TF-IDF value, was used for the TreeTagger morphological analysis. The composition of the Japanese queries of NTCIR1 is as follows.

<TOPIC q=0038>
<TITLE>TCP/IP通信のスループット特性</TITLE>
**<DESCRIPTION>ATM網を用いたTCP/IP通信のスループット特性について述べた論文はないか。**
**</DESCRIPTION>**
<NARRATIVE>新しいネットワーク技術として登場したATM。これをバックボーンとして既存のTCP/IP通信を行なうことができる。ATM網を用いたTCP/IP通信のスループット特性についてのシミュレーションによる評価や解析を行なっている論文が欲しい。ATM網へのTCP/IP通信の接続だけではなく、そのスループット特性についての考察がなければ要求を満たさない。最新の研究動向を知りたい。
</NARRATIVE>
<CONCEPT>
<J.CONCEPT>a. ATMバックボーン, c. スループット, d. 品質評価, e. セル損失</J.CONCEPT>
<E.CONCEPT>a. ATM Backbone, c. Throughput, d. Quality Evaluation, e. Cell Loss</E.CONCEPT>
<A.CONCEPT>a. ATM, b. TCP/IP</A.CONCEPT>
</CONCEPT>
<FIELD>1. 電子・情報・制御</FIELD>
</TOPIC>

<TITLE>: Simple expression of the main concept of the retrieval request.
<DESCRIPTION>: Description of the retrieval request.
<NARRATIVE>: Explanation of the retrieval request, background, details, definition of terms, standard of correct answer judgment, and retrieval purposes.
<CONCEPT>: Presentation of concepts related to the retrieval request, synonym, hypernyms, and hyponyms.

For the experiment, only <DESCRIPTION> was used because it is the most standard retrieval request. As is already known, if the words in <CONCEPT> are added to the retrieval request, the precision increases. However, since the retrieval request most used in information retrieval systems is <DESCRIPTION> in NTCIR1, only <DESCRIPTION> was used in this experiment.
Table 3 shows experimental environment.

Table 3: Experimental environment

| | |
|---|---|
| Computer | Pentium4 2.0GHz, Memory 2GB, HDD 40GB, Fedora Core 4. |
| Language resource | Cross Language WEB-Transer(Machine translation 1), Copyright Cross Language Inc. |
| | KODENSHA J-Server (Machine translation 2), Copyright Kodensha Co., Ltd. With provider Language Infrastructure Group, National Institute of Information and Communications Technology |
| | Online Dictionary of Academic Terms, Copyright National Center for Informatics, Aizawa Laboratory, Digital Content and Media Sciences Research Division, National Institute of Informatics |
| Concept base | Japanese concept base with 200000 words and 98 dimension |
| | English concept base with 100000 words and 151 dimension |

## 5.3 Evaluation

The interpolated recall and precision, the average precision (non-interpolated) for all relevant documents and the precision for 50, 100, 200, 400, 600, 1000 documents were calculated using the TREC evaluation program, which uses the following formula to evaluate the average precision (eq. (5)):

$$\textit{Average Precision} = \frac{1}{D} \sum_{1 \le k \le N} r_k \; x \; \textit{Precision(k)} \quad (5)$$

$$r_k \begin{cases} 1 \text{ (Document of order k represents the correct answer)} \\ 0 \text{ (Document of order k represents incorrect answers)} \end{cases}$$

***D***: number of relevant documents in the retrieval result.
***N***: the relevant document appearing at the end order.
***Precision (k)***: precision at the time of order *k*.

## 5.4 Evaluation results and comparison

For the proposed system, the average precision was evaluated and compared with that of the baseline system and the NTCIR1 participation system. Table 4 shows the evaluation results of the proposed system and the baseline system. The proposed system gives significantly higher precision than does the baseline system. Table 5 shows the evaluation results of the proposed system and the NTCIR1 participation system. The average precision of the proposed system is 0.2142. The best average precision of the NTCIR1 participation system is 0.2109. Again, the proposed system reaches the highest ranking.

Table 6 shows the evaluation results of the proposed system, the query and document translation system of Atsushi Fujii and Tetsuya Ishikawa proposed [6]. They used the same test collection of Japanese-English. Under their system, query and document is translated by MT and human. The average precision values of the proposed system are better than those of the query and document translation system. As for the number of retrieved documents over 200, the average precision values of the query and document translation system that was ideally translated by human are higher than those of the proposed system. As for the number of retrieved documents from 50 to 100, however, the average precision values of the query and document translation system are lower than those of the proposed system. The number of retrieved documents from 50 to 100 is enough for users.

Table 4: Comparison of the results of the proposed system and the Baseline system

| System ID | r-prec | Ave prec |
|---|---|---|
| Baseline system | 0.1753 | 0.1654 |
| **Proposed system** | **0.2396** | **0.2142** |

Table 5: Comparison of the results of the proposed system and the NTCIR1 participation system (Non-interpolated average precision values, averaged over the 39 queries)

| | System ID | r-prec | ave prec |
|---|---|---|---|
| Top 8 Run | BKEBDDS | 0.2225 | 0.2109 |
| | TSB4 | 0.2453 | 0.2090 |
| | TSB3 | 0.2296 | 0.2084 |
| | 1KE3 | 0.2223 | 0.2062 |
| | 1KE | 0.1950 | 0.1940 |
| | 1KE_ij | 0.1976 | 0.1713 |
| | TSB1 | 0.1816 | 0.1617 |
| | TSB2 | 0.1872 | 0.1524 |
| | **Proposed system** | **0.2396** | **0.2142** |
| Monolingual | BKEBDDS | 0.2826 | 0.2618 |

Table 6: Comparison of the results of the proposed system and the query and document translation methods (Non-interpolated average precision values, averaged over the 39 queries)

| Method | | Query and Document translation methods (MT) | Query and Document translation methods (ideal translation by human ) | **Proposed system** |
|---|---|---|---|---|
| Number of retrieved documents (N) | 50 | 0.1690 | 0.1814 | 0.1954 |
| | 100 | 0.1766 | 0.1968 | 0.2058 |
| | 200 | 0.1901 | 0.2142 | 0.2106 |
| | 400 | 0.1946 | 0.2242 | 0.2128 |
| | 600 | 0.1958 | 0.2301 | 0.2137 |
| | 800 | 0.1967 | 0.2319 | 0.2140 |
| | 1000 | 0.1986 | 0.2356 | 0.2142 |

## 6. Discussion

In a query for information retrieval that includes many keywords of relevant documents, a good retrieval result is obtained. If multiple language resources are utilized, the number of keywords of relevant documents is increased in the query. In the proposed method, a query is translated by two machine translation systems. In the query, compound words and noun phrases play important roles in deciding the retrieval result. Therefore, compound words and noun phrases are used for query expansion and for translation by the bilingual dictionary. Other language resources of the two machine translation systems and the bilingual dictionary are not used because the utilization of these other resources does not help to increase the number of keywords of relevant documents, but instead can increase the number of mistranslated words. In this study, the concept base is used to delete mistranslated words from the query. Accordingly, high-precision retrieval results are obtained. That is, the query can be translated with high accuracy. The retrieval performance is improved by re-ranking with the concept base and combining the retrieval results. In this case, the utilization of language resources can enable a query to contain many words that are keywords of relevant documents, and thus, the retrieval results include various relevant documents. However, the language resources also increase the number of words that are keywords of non-relevant documents. Consequently, the rank of relevant documents is decreased and so is the retrieval performance. To increase the retrieval performance, we carry out re-ranking using a concept base with the noun phrases and compound words extracted from the query.

## 7. Conclusions

We set up and implemented a system for improving the performance of cross-language information retrieval by

combining the language resources of the Language Grid. By using multiple machine translation systems, the ambiguity problem was reduced. In addition, by using the bilingual dictionary for translating compound words and noun phrases, the word ambiguity problem was further reduced and the queries were expanded. By implementing a filtering method using the concept base, the mistranslated words in queries were reduced. For information retrieval we used the vector space model with TF-IDF term weighting. The outputs of the top 100 documents were re-ranked by a specifically considered concept base. For the second engine, we used the probabilistic model of the Lemur retrieval system. The final result came from the combined outputs of the two IR engines. Under the proposed method, the cross-language information retrieval system was implemented at a high rate of precision for a NTCIR1 dataset. In comparison with NTCIR1 participation systems, the proposed system attained the highest ranking.

**Acknowledgments**

We received permission to use the test collection of Japanese-English science documents from the NTCIR1 task of the National Institute of Informatics (NII). In addition, we used the language resources of the Language Grid Project during the research process. We would like to take this opportunity to express our sincere thanks to NII and the Language Grid Project.

## References

[1] T. Ishida. “Language Grid: An Infrastructure for Intercultural Collaboration” IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp. 96-100, 2006.
[2] K. Sparck-Jones. “A Statistical interpretation of term specificity and its application in retrieval” Journal of Documentation. vol.28, no.1, pp. 11-21, 1972.
[3] G.Salton. “Introduction to Modern Information Retrieval” McGraw-Hill. 1983.
[4] C. Lin, W. Lin, G Bian, H. Chen “Description of the NTU Japanese-English Cross-Lingual Information Retrieval System for NTCIR Workshop”, NTCIR Workshop 1, 1999.
[5] DW. Oard, J. Wang. “NTCIR CLIR Experiments at the University of Maryland” NTCIR Workshop 1, 1999.
[6]Atsushi Fujii and Tetsuya Ishikawa. “Japanese-English Cross-Language Information Retrieval Integrating Query and Document Translation Methods” IEICE, J84-D-II(2), pp.362-369, 2001.
[7]Aitao Chen et al. “Combining multiple sources for short query translation in Chinese-English cross-language information retrieval.
Proceedings of the fifth international workshop on Information retrieval with Asian languages”, 2000.
[8] Y. Wang, C. Lee, R.Tsai, W.Hsu. “IASL System for NTCIR-6 Korean-Chinese Cross-language information retrieval” NTCIR Workshop 6, pp. 26-30, 2007.
[9] R. Huamg, L. Sun, J. Li, L. Pan, J. Zhang. “ISCAS in CLIR at NTCIR-6: Experiments with MT and PRF” NTCIR Workshop 6, pp. 26-30, 2007.
[10] Atsushi Fujii and Tetsuya Ishikawa. “Japanese/English Cross-language Information Retrieval: Exploration of Query Translation and Transliteration” Computers and the Humanities Vol.35, No.4, pp. 389–420, Nov 2001.
[11] H.Schüetze, J.Pederson. “Information retrieval Based on Word Senses, In Fourth Annual Symposiumon Document Analysis and Information Retrieval”, pp.161-175, 1994.
[12] A.Yasumune, H.Taiichi, T.Takenobu, T.Hozumi. “Research on cross- language information retrieval using vocabulary extension” Natural Language Processing 10, D3-1, 2004..
[13] T.Tokunaga. “Information retrieval and Language processing”, Junichi Tsujii [edit], Foundation the University of Tokyo publication association, Tokyo, 1999.
[14] P. Huy Anh, T. Yukawa. “Cross Language Information Retrieval Based on Concept Base and Language Grid”. ESAIR’10, October 30, 2010, Toronto, Ontario, Canada. ACM 978-1-4503-0372-9/10/2010.

**Pham Huy Anh** received a B.S. in Electrical Engineering from the National Defense Academy and an M.S. in Electrical Engineering from the Nagaoka University of Technology. He is currently a student of doctor course in the Department of Electrical Engineering at the Nagaoka University of Technology.

**Takashi Yukawa** Membership Number of IEEE : [870-2110]
**Takashi Yukawa** received a Master of Engineering degree from the Nagaoka University of Technology in 1987 and a Doctor of Informatics degree from Kyoto University in 2001. He has been involved in the research and development of a parallel computer for expert systems, a concept-sensitive information retrieval system and its application systems, knowledge management systems and an intelligent course management system for e-Learning. He is currently an associate professor in the Department of Electrical Engineering at the Nagaoka University of Technology.

# Enriching Soft Systems Methodology (SSM) With Hermeneutic in e-Government Systems Development Process

**Dana Indra Sensuse[1] and Arief Ramadhan[2]**

**[1] Faculty of Computer Science, University of Indonesia**
**Depok, 16424 , Indonesia**

**[2] Faculty of Computer Science, University of Indonesia**
**Depok, 16424 , Indonesia**

**Abstract**

e-Government system has been developed in various countries. e-Government system can support the government's performance in serving the public. Because there are many aspects that must be considered, then the e-Government system development process can be very complex. Soft Systems Methodology (SSM), that is based on soft system thinking, is suitable for use in the e-Government system development process. Hermeneutic can be used to enrich SSM. Hermeneutic can be used to uncover knowledge, spesifically from the interview or Focus Group Discussion (FGD) result. Hermeneutic can be done using hermeneutic circle principle. Hermeneutic can be used in the first step, the second step and the sixth step of SSM.

***Keywords:*** *e-Government, Soft Systems Methodology, Hermeneutic*

## 1. Introduction

Currently the development of e-Government systems have been proliferated in several countries, both in developing countries and developed countries. The development of e-Government system can support the government's performance in serving the public.

e-Government is the use of Information Technology (IT) by public sector organizations [1]. Other definition of e-Government is public sector use of the Internet and other digital devices to deliver services, information, and democracy itself [2].

e-Government is closely related to computer, information and cyber world. Thus, Ramadhan, Sensuse, and Arymurthy said in [42] that computer ethics, information ethics, and cyber ethics are the foundation for e-Government ethics.

The main orientation of e-Government is the accessibility of information by the public [1]. Heeks in [1] says that e-Government is also an information system, but it is enriched with various aspects, such as the management aspects, political aspects, economical aspects and others. These aspects have to be considered by developers when developing an e-Government system.

Because many aspects that must be considered, then the e-Government system development process can be very complex. These aspects can not be observed separately, but should be observed as a whole, where there is interaction in it. Such characteristics can be solved using systems thinking.

The approach of systems thinking is fundamentally different from the traditional form of analysis [3]. Instead of focusing on the individual pieces of what is being studied, systems thinking focuses on the feed-back relationship between the thing being studied and the other parts of system [3]. Therefore instead of isolating smaller parts of a system, systems thinking involves a broader view, looking a larger numbers of interactions [3]. In this way, systems thinking creates a better understanding of the big picture [3]. Based on the above explanation, it can be said that the systems thinking is more holism and non-reductionism.

There are two forms of systems thinking, i.e. hard systems thinking and soft systems thinking. Hard systems thinking assumes the world is mechanic, contains systems which can be modeled and “engineered” [4]. Hard systems thinkers assume reality to be objective, that reality looks the same regardless of who is the observer [4].
Soft systems thinking do not assume that the world is systemic and well-ordered; on the contrary, it assumes social reality to be “problematical”, characterized by multiple angles of approaches and perspectives [4]. The understanding of reality is dependent upon the observer,

his interpretations and what he chooses to focus on [4]. This suggests that soft systems thinking is more subjective than objective.

e-Government is a socio-technical system that consists of soft components and hard components [1]. It could be argued that the soft component is the people who are involved in e-Government, whereas the hard component is the Information Technology (IT) that being used. The management approach of the soft component is likely inspired by social sciences, it tends to be subjective, qualitative, and further highlight by the aspects of humanism [1].

Because of the existing of soft component in e-Government that tends to be subjective as stated in [1], then we consider that soft system thinking is suitable for use in the e-Government system development process. Some of the methodologies that can be used in soft systems thinking is a meta-synthesis approach as used in [5], and Soft Systems Methodology (SSM). From these two options, we chose SSM as a basis in our discussion of this paper. In addition, we also will explain that hermeneutic can be used to enrich the methods that already exist in the SSM.

## 2. Enriching Soft System Methodology

Soft Systems Methodology (SSM) was proposed in 1981 by Peter Checkland [6]. As the name implies, SSM is based on soft systems thinking. The picture of SSM can be seen in Fig. 1.

SSM consists of seven steps, i.e (extracted from [6]):

1) The identification of a problem situation that demands attention
2) Problem situation is expressed. The expression can be described using the Rich Picture Diagram. The examples of the Rich Picture Diagram can be seen in Fig. 2.
3) Some relevant human activity systems, potentially offering insight into the problem situation, are selected and from these 'root definitions' are built. In this step, CATWOE analysis is performed. CATWOE stands for Customers, Actors, Transformation process, World view, Owner, and Environmental constraints.

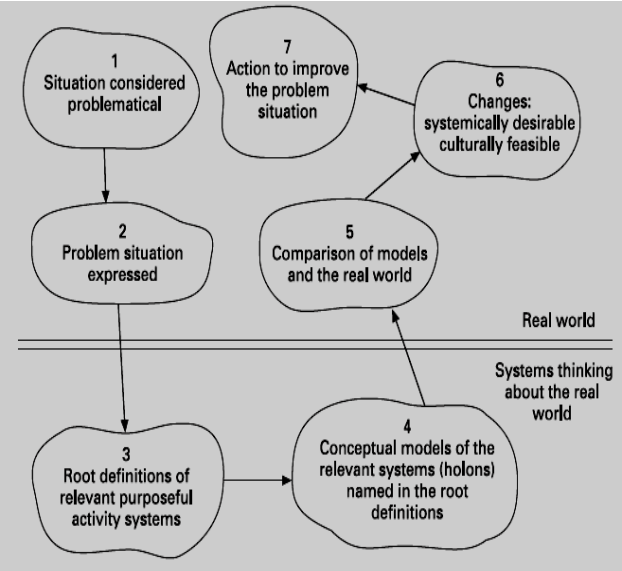



Figure. 1. The seven-step of Soft System Methodology (SSM) [6].

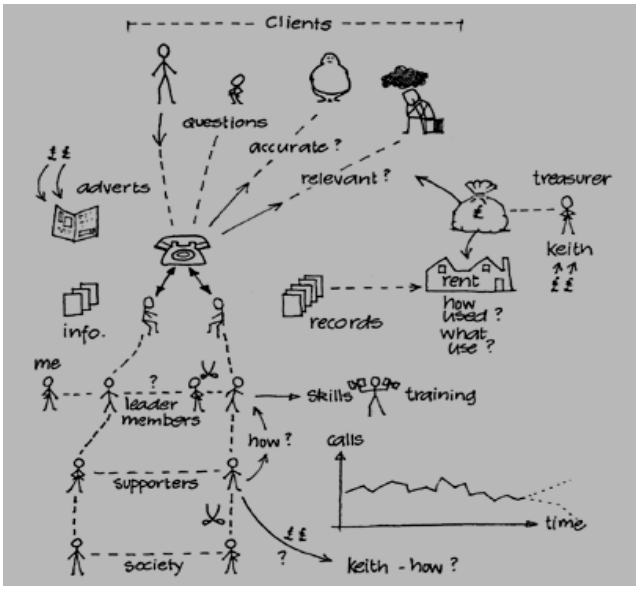



Figure. 2. Part of a rich picture of a telephone helpline situation[7].

4) Construct conceptual models. This is the most important step in the SSM. Various modes of modeling techniques can be applied at this step.
5) Comparing the conceptual model with the real world. The aim is to provide material for debate about possible change among those interested in the problem situation. This step shows the social processes within the SSM.
6) Making changes to the model by accommodating the interests of several actors involved. Changes should be able to follow the desired model but still possible (feasible) historically, culturally and politically. Changes may include changes in attitudes, structures,

or procedures.

7) Perform various activities to implement the model and fix the problem. In this step, the conclusions are drawn and long-term solution is formulated.

SSM has been amended several times. The first change is made in 1990 in the form of "two-strands model" as shown in [6]. In this model, were added three types of inquiry, referred to as Analysis 1, 2 and 3 [6]. Analysis 1 considers the intervention itself and the roles of client, problem-solver and problem-owners. Analysis 2 is social system analysis. Analysis 3 examines the politics of the problem situation and how power is obtained and used [6].

Subsequent changes of SSM is made when the original seven-step is merged into just four steps [8]. The new four-step is named as “learning cycle of SSM”. Four new steps are [8]:

1) Finding out about a problem situation, including culturally/politically.
2) Formulating some relevant purposeful activity models
3) Debating the situation, using the models, seeking from that debate both :
   a. changes which would improve the situation and are regarded as both desirable and (culturally) feasible
   b. the accommodations between conflicting interests which will enable action-to-improve to be taken
4) Taking action in the situation to bring about improvement.

Although the SSM has been amended several times and although Checkland no longer favours it, the representation of SSM as a seven-step, which appeared in 1981, is still frequently used today [6]. Some examples of the use of the seven-step of SSM can be seen in the [9], [10], [11], [12] and [13].

Yang & He use the seven-step of SSM for regional planning and improvement of industrial structure in Guangdong, China [9]. Jianmei & Zheng use the seven-step of SSM to analyze the soft conflict of interest problem [10]. Lehaney & Paul use the seven-step of SSM in simulation modeling [11]. Dirker *et. al*. in [12], using a seven-step SSM to manage the distribution channels in a production market. Kang & Hu in [13], using a seven-step SSM within the logistics system.

Besides just using it, some researchers have also tried to adding a new method to the SSM or mixing SSM with others. Biggam equipped the seven-step of SSM with knowledge type, to facilitate the process of knowledge capture in the website environment [14]. Yinghong in [15], complete the seven-step of SSM with integrated decision making knowledge system in decision-making process. Razali *et. al*. in [16] integrate the seven-step SSM into the Design Science Research (DSR) for modeling a system framework. And the new one, Ramadhan, Sensuse and Arymurthy in [17], adding the Focus Group Discussion (FGD) method in the step two and the step six of the seven-step of SSM.

In this paper we try to enrich the SSM with the hermeneutic. We suggest that with the addition of hermeneutic into the SSM, then the process of uncover knowledge, in the e-Government system development process, can be done better.

## 3. Hermeneutic and Hermeneutic Circle

Hermeneutic generally used in interpretive research. Hermeneutics can be considered as a theory or philosophy of the interpretation of meaning [18]. Not just a teory, hermeneutic also is the study of practice in understanding and interpretation.

The name hermeneutics is associated with “Hermes”, the Greek god of communication, the borders, the limits [19]. It represents the crossing of paths and the coincidence of moments [19]. The term ‘hermeneutic’ derives from Greek language for ‘to interpret’ or ‘to understand’ [20].

Hermeneutics is primarily concerned with the meaning of text [21]. Hermeneutic initially focusing on the interpretation of sacred texts and law. Since the main object is text, then hermeneutic closely associated with reading and therefore also to literacy [22].

The use of hermeneutics has grown from its roots in the interpretation of Greek classical literature [27]. Currently, hermeneutic also be used in making the interpretation of texts in other disciplines, for example in the field of Computer Science as being practiced in [23], [24], [25], and [26].

The key word in hermeneutics is interpretation, as stated in [28]. The main assumption is that every human being perceives herself and her situation in a special way, and that she applies meaning to everything that surrounds her [28]. This means science that aims to find out something about something involves interpreting the world and how it is perceived, not how it is [28].

Hermeneutics presents us with a way of exploring

meaning and interpretation that incorporates several key aspects: a relational approach to the text (information, event); the circular and evolutionary character of dialogical understanding; and the importance of prior understandings, which have a historical dimension and stress the contextuality of meaning from a pragmatic recognition of the world and our place in it with others [22].

Five characteristic of Hermeneutic are [29] : (a) seeks understanding rather than explanation; (b) acknowledges the situated location of interpretation; (c) recognizes the role of language and historicity in interpretation; (d) views inquiry as conversation; and (e) is comfortable with ambiguity.

Hermeneutic closely related to postmodernism. This is inline with what is expressed in [30], that the hermeneutic concept of the interpretive circle suggests that interpretation is an endless process while postmodern perspectives similarly emphasize the multiplicity and infinity of interpretations.

Ramadhan, Sensuse and Arymurthy has been revealed in [31], that one of the characteristics of postmodernism is all meaning is contextual and based on difference. That characteristic appear in the hermeneutic. As stated in [32], that a hermeneutic researcher approaches his research object with his own understanding as an advantage. The underlying problem is that every reader of a text has a different understanding of that text depending on his or her own experiences and life-world [33].

Other characteristic of postmodernism is the existence of social construct [31]. That characteristic also appear in hermeneutic.

In [31], it was argued that postmodernism can appear in e-Government system development process. Then, based on some of the explanations above, it is known that some characteristics of postmodernism is also present in hermeneutic. This suggests that the hermeneutic can also involved in e-Government system development process.

Hermeneutic interpretation performed with a fundamental principle called the hermeneutic circle [34]. In hermeneutic circle principle, interpretations made from a specific section ("part") and then moves to the common section ("whole"), then return from ("whole") toward ("part") [34]. This is done iteratively in order to get a thorough understanding [34].

Examples of the repeated interpretation process between whole and part as a hermeneutic circle can be found in [35], as shown in Fig. 3. The middle circle in Fig. 3 shows the hermeneutic interpretive process as alternating between interpretations of observations of community meetings (minutes of meetings and field notes), public documents (reports and website material disseminated by local authorities) and interviews (interview transcripts) [35].

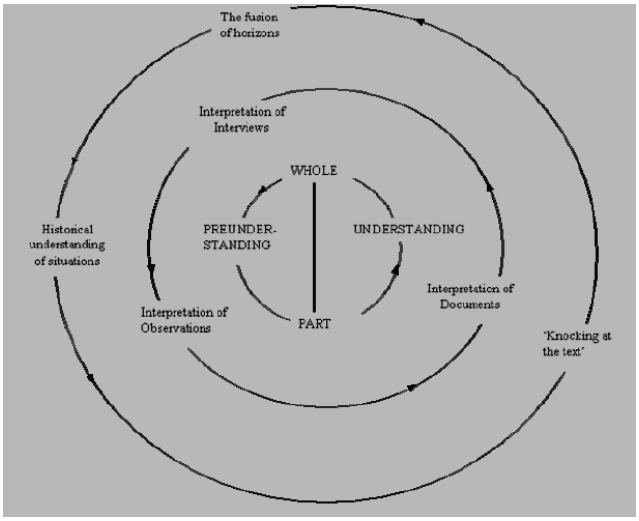


Figure. 3. Example of hermeneutic circle application [35].

The hermeneutic circle, therefore, allows the researcher to uncover in-depth and rich data [24]. The idea behind hermeneutic circle is that there is a circular relationship between the prior knowledge of a recipient of a text and his or her understanding of the same text [33]. By the hermeneutic circle, it means that hermeneutic can be done iteratively as implied in [26].

Hermeneutics has moved away from the idea of such a 'correct' understanding and has expanded into the art of understanding all communication, not just written text [33]. As revealed in [30], hermeneutic can also be used to observe the social practices. Hermeneutic can also be done to interpreting the meaning of experiences and symbolic artefacts (such as art or sculpture or architecture), which may be either historical or contemporary [36].

## 4. Where And How Hermeneutic Can Be Used in SSM

In the use of SSM, a developer of e-Government system can use the interview as one of his methods. The interviews can be done in the first step of SSM, i.e. the identification of a problem situation. Objects that can be

interviewed are the stakeholders of e-Government system that will be developed. The interview results are then written in the form of text-based transcripts.

In addition, to support the first step of SSM, a developer is sometimes necessary to collect some related documents. These documents can be in the form of law document, government agency's strategic plan document, current infrastructure condition document, and others. Again, these documents are of course in the form of text.

To support the process of developer understanding about some text-based resources above (i.e. the interview transcript or others document), and to uncover knowledge from them, then the developer can do hermeneutic. This is consistent with what is stated in [37] that hermeneutic primary concern is with the meaning of a text or text-analogue that can be (for example) a book, scholarly article, interview transcript, email, or organization.

Hermeneutic can be done using hermeneutic circle principle. Transcriptions of interview data were repeatedly searched for themes and ideas [38]. The study iterated between the fragments of interviews as parts and the global context [38].

Our suggestion is also supported by a statement that is stated in [39], that hermeneutic can be used to uncover knowledge. From hermeneutics we can learn that the process of producing knowledge has to do with the creation of texts and that the reception and development of texts are what constitutes knowledge [33]. Specifically, hermeneutic can be used to uncovered the tacit knowledge as being implied in [40]. It is also revealed in [41], that hermeneutics were utilised because they enabled the understanding of the impact of meta-abilities on tacit knowledge externalisation and sharing from the text (text-analogue).

To facilitate the developer, currently hermeneutic can also be conducted computerized. One software example that can be used to do hermeneutic is NVIVO.

If the developers are also use the FGD method in the SSM, such as those proposed in [17], then the hermeneutic can also be performed. In the [17], it was argued that the FGD can be done in the second step and sixth step of SSM. The results of the FGD then being recorded or written in the narrative form by the moderators. In this case, hermeneutic can be performed on that narrative text.

From the above explanation it can be concluded that the hermeneutic can be used in the first step of SSM, i.e. the identification of a problem situation. In addition, if the developer is also doing FGD as being proposed in [17], then the hermeneutic also can be done in the second step and sixth step of SSM.

## 5. Conclusions

Since there are a lot of aspect involved and there exist soft component in e-Government, then the e-Government system development process can be done using the Soft System Methodology (SSM), which is based on soft systems thinking. To uncover knowledge, then the hermeneutic can be used to enrich the SSM. Hermeneutic can be done using hermeneutic circle principle. The hermeneutic can be used in the first step, second step and sixth step of SSM.

## References

[1] R. Heeks, Implementing and Managing eGovernment An International Text, London, England : SAGE Publications, 2006.
[2] D. M. West, Digital Government Technology and Public Sector Performance, New Jersey, USA : Princeton University Press, 2005.
[3] D. Aronson. (1999). Targeted Innovation Using Systems Thinking to Increase the Benefits of Innovation Efforts [Online] Available: www.thinking.net/Systems_Thinking/st_innovation_990401.pdf.
[4] A. Mirijamdotter, "A Multi-Modal Systems Extension to Soft System Methodology," Ph. D. dissertation, Lulea Tekniska Universitet, Sweden, 1998.
[5] J. Gu and X. Tang, "Meta-Synthesis System Approach To Knowledge Science," International Journal of Information Technology & Decision Making, vol. 6, no. 3, pp. 559-572, 2007.
[6] M. C. Jackson, System Thinking Creative Holism for Managers, John Wiley & Sons Ltd, England, 2003.
[7] http://systems.open.ac.uk/materials/T552/pages/rich/richAppendix.html
[8] P. Checkland, "Soft Systems Methodology: A Thirty Year Retrospective," Systems Research and Behavioral Science, vol. 17, pp. S11-S58, 2000.
[9] J. Yang and Y. He, "An Application of Checkland's Soft System Methodology in China", in Proceedings of IEEE International Conference on Systems, Man and Cybernetics, IEEE Press, vol. 1, 1995, pp. 603-608.
[10] Y. Jianmei and H. Zheng, "A Suggestion of SSM with Interest-Coordination Process", in Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, IEEE Press, Vol. 3, 1996, pp. 2412-2416.
[11] B. Lehaney and R. J. Paul, "Soft System Methodology and Simulation Modelling", in Proceedings of the 1996 Winter Simulation Conference, IEEE Press, 1996, pp. 695-700.
[12] H. G. Dirker et. al., "A Systems Thinking Approach to Manage Distribution Channels in the Control and Instrumentation Product Market with Multi-Faceted Product

Lines,"in Proceedings of the 2008 IEEE ICMIT, IEEE Press, pp. 1389-1394.
[13] B. Kang and J. L. Hu, "Research and Improvement of the Logistics System Based on Soft Systems Methodology," in Proceedings of IEEE International Conference on Advanced Management Science (ICAMS), IEEE Press, vol. 1, pp. 117-120, 2010.
[14] J. Biggam, "Exploiting Soft Systems Methodology (SSM) and Knowledge Types to Facilitate Knowledge Capture Issues in a Web Site Environment", in Proceedings of the 35th Hawaii International Conference on System Sciences, IEEE Press, 2002.
[15] Z. Yinghong, "Soft Systems Methodology Based on Decision Making Knowledge Integration," in Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing 2007 (WiCom 2007), IEEE Press, pp. 5733-5736, 2007.
[16] S. Razali et. al., "Applying Soft System Methodology (SSM) into the Design Science: Conceptual Modeling of Community based E-museum (ComE) Framework," in Proceedings of IEEE International Conference on Systems Man and Cybernetics (SMC), IEEE Press, pp. 2701-2707, 2010.
[17] A. Ramadhan, D. I. Sensuse, and A. M. Arymurthy, "A proposed methodology to develop an e-Government system based on Soft Systems Methodology (SSM) and Focus Group Discussion (FGD)," in Proceedings of 2011 International Conference of Advanced Computer Science and Information Systems (ICACSIS), IEEE Xplore, 2011.
[18] C. D. Pedron and A. Z. Saccol, "What Lies behind the Concept of Customer Relationship Management? Discussing the Essence of CRM through a Phenomenological Approach," Brazilian Administration Review, vol. 6, no. 1, 2009, pp. 34-39.
[19] J. Barojas, "Problem solving and writing II: The point of view of hermeneutics", Lat. Am. J. Phys. Educ, Vol. 2, No. 1, 2008, pp. 6-14.
[20] D. Hart, "Systemic Evaluation Methodology For Technology Supported Learning", Ph.D. Dissertation, The University of Sheffield, 2010.
[21] K. D. Peszynski, "Power and Politics in a System Implementation", Ph.D. Dissertation, Deakin University, 2005.
[22] R. T. O'Farrill, "Information Literacy and Knowledge Management: Preparations for an Arranged Marriage", Libri, vol. 58, 2008, pp. 155–171.
[23] S. R. B. Berdal, "Public deliberation on the Web: A Habermasian inquiry into online discourse", Cand.Scient Thesis, University of Oslo, Oslo, Sweden, 2004.
[24] J. K. B. Yeo, "An Investigation of Contextual Factors Influencing The Development of a Sustainable Knowledge Economy", Ph. D Dissertation, The Pennsylvania State University, Pennsylvania, USA, 2007.
[25] T. Butler and C. Murphy, "Understanding the design of information technologies for knowledge management in organizations: a pragmatic perspective", Info System Journal, Vol. 17, 2007, pp. 143-163.
[26] J. D. Warren, "TV in The Age of The Internet: Information Quality of Science Fiction TV Fansites", Ph. D Dissertation, Indiana University, Indiana, USA, 2011.
[27] C. von Zweck, M. Paterson, and W. Pentland, "The Use of Hermeneutics in a Mixed Methods Design", The Qualitative Report, Vol. 13, No. 1, March 2008, pp. 116-134.
[28] M. Runardotter, "Information Technology, Archives and Archivists – An Interacting Trinity for Long-term Digital Preservation", Licentiate Thesis, Lulea University of Technology, 2007.
[29] E. A. Kinsella, "Hermeneutics and Critical Hermeneutics: Exploring Possibilities Within the Art of Interpretation", Forum: Qualitative Social Research, Vol. 7, No. 3, Art. 10, May 2006.
[30] V. Mottier, "The Interpretive Turn: History, Memory, and Storage in Qualitative Research", Forum: Qualitative Social Research, Vol. 6, No. 2, Art. 33, May 2005.
[31] A. Ramadhan, D. I. Sensuse, and A. M. Arymurthy, "Postmodernism in e-Government", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No. 1, July 2011, pp. 623-629.
[32] J. Jonsson, "Strategic understanding of a complex customer system A field experiment at UNOSAT", Master's Thesis in Informatics, Goteborg University, Sweden, 2004.
[33] B. C. Stahl, Information Systems Critical perspectives, London, England : Taylor & Francis, 2008.
[34] H. K. Klein and M. D. Myers, "Evaluating Interpretive Field Studies In Information Systems", MIS Quarterly, Vol. 23, No. 1, 1999, pp. 67-94.
[35] M. Arunachalam, "A Philosophical Hermeneutics Approach for Understanding Community Dialogue on Environmental Problems: A Case Study of Lake Taupo," in Proceedings of The 5th European Conference on Research Methodology, pp. 31-39.
[36] J. Gupta, "Potential factors influencing adoption of a Service Oriented Architecture: Experiences from specialist healthcare in Norway", Master Thesis, University of Agder, 2008.
[37] J. Fowler, P. Horan, and C. Cope, "How an "Imperative" IS Development was Saved from a Failing Course of Action – A Case Study", Issues in Informing Science and Information Technology, Vol. 4, 2007, pp. 395-406.
[38] T. Linden and J. L. Cybulski, "Application of Grounded Theory to Exploring Multimedia Design Practices," in Proceedings of 7th Pacific Asia Conference on Information Systems, 2003, pp. 508-522.
[39] H. Klaus, "Elements of a Hermeneutics of Knowledge in Government The Coalition of Public Sector Reform and Enterprise Resource Planning", Ph.D. Dissertation, Queensland University of Technology, 2004.
[40] P. Busch and D. Richards, "Graphically defining articulable tacit knowledge", in Proceedings of Conferences in Research and Practice in Information Technology, Vol. 2. 2001.
[41] M. H. Selamat, "Developing Individuals for Developing Learning-Based Systems", Ph.D. Dissertation, Brunel University, 2005.
[42] A. Ramadhan, D. I. Sensuse, A. M. Arymurthy, "e-Government Ethics : a Synergy of Computer Ethics, Information Ethics, and Cyber Ethics", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No. 8, 2011, pp. 82-86.

**Dana Indra Sensuse** B.Sc in Soil Science (Bogor Agricultural University, Indonesia, 1985), M.Sc in Library and Information Studies (Dalhousie University, Halifax, Canada, 1994), Ph.D. in Information Studies (Toronto University, Canada, 2004), Lecturer at University of Indonesia, Head of e-Government Lab at University of Indonesia.

**Arief Ramadhan** B.Sc in Computer Science (Bogor Agricultural University, Indonesia, 2005), M.Sc in Computer Science (Bogor Agricultural University, Indonesia, 2010), Ph.D. Student in Computer Science (University of Indonesia), Research Assistant at University of Indonesia. Member of e-Government Lab at University of Indonesia.

# Reuse of Use Cases Diagrams: An Approach based on Ontologies and Semantic Web Technologies

**Belén Bonilla-Morales[1], Sérgio Crespo[2] and Clifton Clunie[3]**

**[1] Universidad Tecnológica de Panamá**
**Panama City, Panama**

**[2] Universidade do Vale do Rio Dos Sinos**
**Sao Leopoldo, Brasil**

**[3] Universidad Tecnológica de Panamá**
**Panama City, Panama**

**Abstract**

Software reuse is defined as the use of any artifact, or part thereof, created before, on a new Project. This practice has significant benefits in reducing costs and increasing quality and productivity in software development. Numerous approaches have been proposed aimed mostly at the source code reuse, but this type of reuse has its limitations because development platforms and technologies are constantly changing. Then, it is necessary to apply reuse over software artifacts created at higher levels of software life cycle such as requirements specification. This paper presents a tool for the reuse of use case diagrams by storing their information in OWL ontology and the use of Semantic Web technologies.

***Keywords:*** *Software reuse, use cases diagram, ontology, Semantic Web.*

## 1. Introduction

Software reuse is defined as the process of creating software systems from existing software; it is the use of any device or part thereof, previously created in a new project [1].

Reuse of software has been the subject of research for many years in the software community because of the great benefits it provides, mainly in terms of reducing development time and cost, and increasing the quality of software systems [2].

In most cases, software reuse is associated only with source code reuse, but as indicated in its definition, we can reuse any type of software artifact [3]. Furthermore, source code reuse turns out to be problematic because the development platforms and technologies are constantly changing. Therefore, it is necessary and appropriate to apply software reuse over artifacts created at higher levels of software life cycle such as requirements specification which includes use case diagrams.

Use cases diagrams are a type of UML diagram whose purpose is to define graphically the functionality of a system in terms of actors, use cases and relations [4]. They have great importance as a technique for extracting and defining functional requirements from the user point of view [5].

Reuse of use cases diagrams then gives the opportunity to have formal definitions and validated functional requirements of a software system created earlier and thus be able to use them as often as necessary in different software projects. In addition, reuse could be applied on a higher level of abstraction, avoiding the limitations in terms of changes in technologies and platforms.

But it is not enough with the initiative and the definition of the process if we do not have the necessary tools to operationalize the idea of reuse of requirements specifications. Consequently, we implemented a tool that allows reuse of use case diagrams by storing, searching and retrieving them using ontologies and Semantic Web technologies. The application allows software engineers to store the information of use case diagrams into OWL ontology. Then, we can make semantic searches on the repository where we store the RDF triples that define the ontology. Thus, the software development teams can create their ontological database of use cases diagrams and to have an application that allows to search and retrieval them, and finally reuse them in new projects. It promotes a savings of time and effort during the stage of requirements analysis and modeling of software.

## 2. State of Art

Reuse of software is generally defined as the process of building or assembling software applications or software systems from previously developed software [1] [2]. Its application involves not only the source code, but also the different artifacts that occur during the life cycle of software like requirements, UML diagrams, tests, manuals, experiences, etc. [3].

The benefits gained through the reuse of software artifacts are many, but certainly the most important are:

- A reduction in development costs.
- An increase in product quality.
- An increase in productivity through improvement of the times in which it is developed new software projects.

Reuse of software has been a topic of popular debate and research for nearly forty years in the software community. Many approaches to software reuse have been proposed during this time, mainly oriented to the development of tools and methodologies.

Among the outstanding works we can mention the proposed by Monegan [6] who developed a tool called Object-oriented Reuse Tool (ORT) which supports the reuse of object-oriented software by maintaining a library of reusable classes and record information about reuse and information associated with design and verification. For his part, Gicca [7] developed a software tool in ADA language called Reuse System to promote the reuse of software components and requirements, high level design, source code, etc. through a repository. Unlike the tool proposed by Monegan, Gicca's tool supports reuse of software components in the different phases of software life cycle, not just source code.

On the other hand, Henniger [8] developed the tool CodeFinder which is composed of three main parts: the tool (PEEL) which is responsible for populating the repository with reusable components of functions and routines obtained from source code in Emacs Lisp [9], a search mechanism, and a fitting tool to refine the repository when necessary. The tool uses two techniques: an intelligent recovery method which finds information related to the query, and a query building supported through incremental refinement of queries. The system provides a user interface that implements the search and navigation mechanisms that allows the user to view and navigate the hierarchy of the repository and build search queries.

There has been more specific work focused on the reuse of UML diagrams, such as Blok and Cybulski [10] who proposed a method for reuse use cases specifications using WordNet language to classify the lexical and semantic flows events. The tool calculates the similarity of the use cases according to information obtained in the flow of events and uses an information retrieval technique. Meanwhile, Robinson and Woo [11] proposed techniques for reuse UML artifacts, specifically the sequence diagrams. The main idea of the work is to find the model that best fits the desired features or functionality through REUSE tool, which uses Subdue algorithm [12] to find links between different sequence diagrams using the information stored in elements of stereotypes such as names, classes, etc. These two works differ from ours in aspects like UML software artifacts to reuse and the methodology used.

Happel et al. [13] presented Kontor, an approach which aims to store and query XML-based metadata, on different software artifacts, including UML components, in a central repository to encourage reuse. It has several ontologies to describe knowledge about the artifacts and technologies and / or programming languages, software licenses, etc. The work also includes a number of SPARQL queries that can be executed by the software developer to recover pieces of software to fit a need to develop specific application. In this work, like in ours, ontologies are used to store and retrieve software artifacts, but it is not geared to specific reuse of use cases diagrams.

## 3. Reuse of Use Cases Diagrams: An Approach based on Ontologies and Semantic Web Technologies

As part of our research on the topic of software reuse, we developed a tool that allows reuse of use cases diagrams in UML. The main idea of this paper is to manipulate and store relevant information of use cases diagrams in an OWL ontology. By having this ontological representation of the use cases diagrams information within a repository, the tool allows the user to make parameterized queries, through a graphical interface, about the diagrams that he is interested in getting, while it increasing the accuracy of the results obtained by taxonomic characteristics, capacity for inference and management concepts that have OWL ontologies.

Below, it is presented the technologies used to develop the tool. Then, we define the architecture of the tool and its implementation details.

### 3.1 Technologies and Tools

In this section we briefly describe the technologies and software tools used during the development of our tool:

a) UML and Use Cases Diagrams

UML is the most used and known language to model application structure, behavior and architecture but also business process and data structure. It is a graphical

language for visualizing, specifying, constructing, and documenting a software system [14].
Use cases diagrams are a type of UML diagram. Use cases diagrams are a technique to specify the behavior of a system, capture its requirements and guide the development process [15].
The most important concepts that define the use cases diagrams are:
**Actors**: An actor is a representation of a person, system or machine that interacts with the software system being developed.
**Use Case**: A task that must be undertaken with the support of the system being developed. A use case represents a particular functionality of the system.
**Relations**: Represent dependency between use cases or between actors and use cases, so that we can set the behavior of the system by integrating each feature. Use case diagrams in UML support partnership, inclusion, extension and generalization relations.

b) XML Metadata Interchange – XMI

XMI is a standard of the Object Management Group (OMG) for exchanging metadata level information via XML. It is a specification for exchanging diagrams.
According to the active site OMG, XMI is defined as: "A model driven XML integration framework for defining, interchanging, manipulating and integrating XML data and objects" [16].
XMI architecture allows simplifying the communication between applications of different technologies saving much time and works, also enhances the reuse of objects and components.

c) Ontology Web Language – OWL

OWL is a markup language for publishing and sharing data using ontologies, which are defined as explicit specifications of conceptualizations, in order to enable the behavior and reuse of knowledge [17] [18].
OWL is a Semantic Web technology, which is defined by the W3C as an extended Web, endowed with greater meaning in which any Internet user can find answers to his questions more quickly and easily with better-defined information. Thus, through the Semantic Web we can get solutions to common problems in finding information through the use of a common infrastructure, whereby it is possible to share, to process and transfer information easily [19].

d) Java and Netbeans IDE

Java is an object-oriented programming language developed by Sun Microsystems, independent of the platform on which we use the applications made with this language. For its part, Netbeans is a development platform that lets us work with Java and other programming languages. It lets work with the Swing graphics library.

e) JDOM

JDOM is an open source library for handling XML data optimized for Java. It allows create, read and manipulate XML files easily in any Java application [20].

f) Jena Framework

Jena is a Java framework for building Semantic Web applications. It provides a collection of tools and Java libraries to help us to develop semantic web and linked-data applications, tools and servers. It includes a programming environment for RDF, RDFS and OWL, and persistent memory storage, a SPARQL query engine and an inference engine based on rules [21].

g) SPARQL Language

SPARQL is a query language and a protocol for accessing RDF. Like SQL, it is necessary to distinguish between the query language and the engine for storage and retrieval of data; for this reason, there are multiple implementations of SPARQL, usually associated with development environments and platform technologies. For our specific case, ARQ is used as a query engine that supports SPARQL [22].

## 3.2 Tool Architecture

The steps required to implement reuse of use case diagrams through our tool is defined as follows:
**Creation of use case diagrams and exportation to XMI:** The use cases diagrams are modeled using a CASE tool for UML like StarUML [22], Rational Rose [23], ArgoUML [24], among others. Once they are created, we export them via XMI export option offered by most of these programs. The XMI files are stored in a specific directory within the server.
**Storage of information of use case diagrams into OWL ontology:** Our tool has an option that allows the user to load XMI files derived from use cases diagrams in UML. Also it has a GUI that allows entering information which can enrich the knowledge we have of the diagrams. Immediately after the XMI file is loaded into the system, tool handles it internally to obtain information about actors and use cases. Once this information and the information entered through the GUI are obtained, it creates new individuals which form part of the OWL ontology.
**Persistent storage of OWL ontology in MySQL:** Once the new individuals are obtained, the ontology is stored in a MySQL database. Thus, the ontology and its instances will be available for later reference.
**Search and retrieval use case diagrams in XMI format:** Through the tool, the user can search use case diagrams

that were stored previously. The user employs a number of parameters to define his query, the tool interprets the user's request as a query over OWL ontology that is stored in MySQL, then it retrieves those individuals that match the query and finally results are presented as XMI files associated with these individuals or entities.
Figure 1 shows a component diagram that conceptualizes the architecture of our tool, according to the phases we have described here.

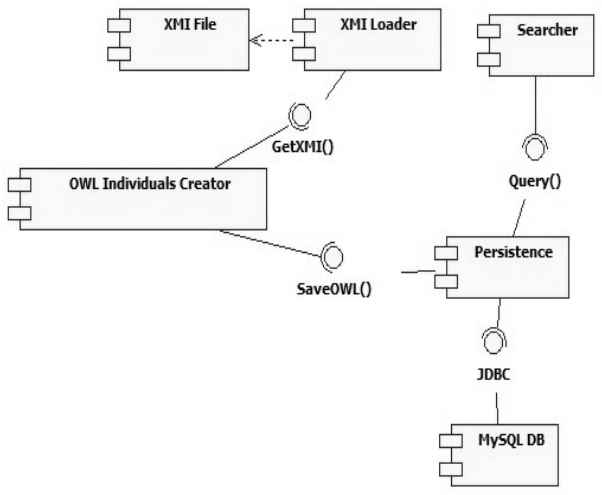



Fig. 1 Architecture of the tool

## 3.3. Implementation

Our tool is developed using Java programming language and the Jena framework. It consists of two main GUIs created using the Swing graphics library: one for the XMI file upload and insert additional information for use cases diagrams and another for searching and retrieving information.
It was created an OWL ontology, using the Jena framework, which has all the necessary classes and properties to store information of diagrams. The ontology provides a categorization by type of business and project.
Figure 2 shows the graphical interface through which the user uploads an XMI file and enters additional and relevant information associated with the use cases diagram. The user searches and enters the XMI file and fills additional information about the use case diagram. A class named XMILoad handles the XMI file loading process and obtains actors and use cases of the file with the help of JDOM, and the class named AddIndividuals builds the OWL structures necessary to add new individuals with their properties in the base ontology.

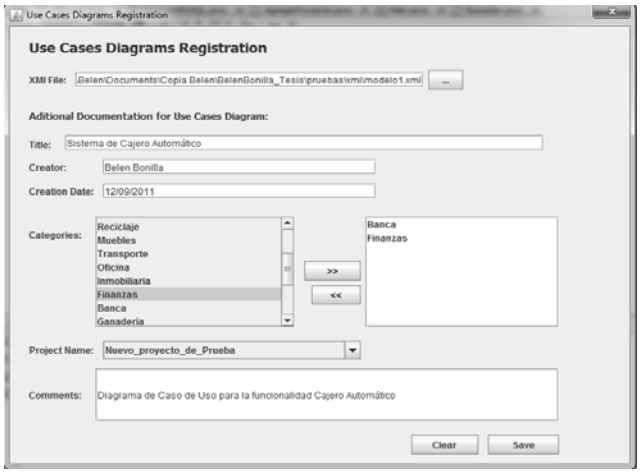



Fig. 2 GUI for Use Cases Diagrams Registration

The base ontology is stored in a MySQL database. Jena allows persistent storage of ontologies through its RDB subsystem. A class named AddIndividuals makes a connection to the base ontology stored in MySQL and adds any new individuals.
As part of the additional information, it is allowed to enter the category or type of business which is associated with the use case diagram, besides the project to which it belongs. This information is available because there is an interface that allows adds new categories and another to enter new projects. Then, these new categories and projects are also stored in the ontology as individuals thus allowing being loaded from the database to their use.
The search and retrieval of the use case diagrams in XMI format is carried out through two classes, one for the building of SPARQL queries and another for the retrieval and presentation of information.
Figure 3 shows the graphical interface to search use case diagrams in XMI format.

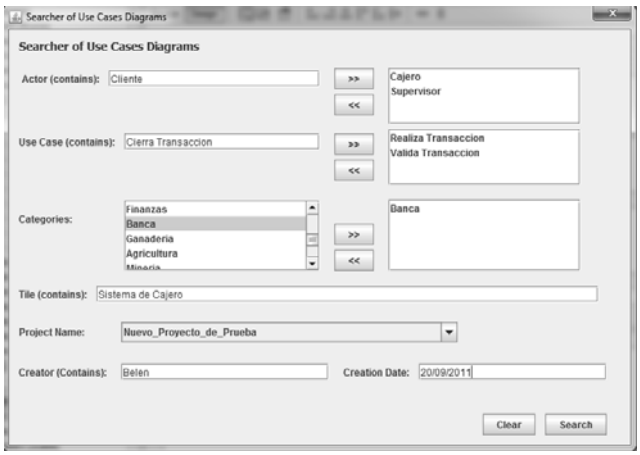



Fig. 3 GUI for Search Use Cases Diagrams

Through a class named GetQuery it is constructed SPARQL sentences from the information entered by the user via the GUI.
A class, named IndividualsRecovery, brings back and presents use case diagrams in XMI format which match with user search's parameters, including a brief description of them. This is possible because individuals in the ontology, that represent use case diagrams, have an ID which identifies them with their respective XMI file. The user will select the XMI file he wants and will download it for reuse.
Each time an XMI file that defines a use case diagram is downloaded for reuse, a weight will be added in the ontology so that it serves as feedback to determine how useful being the diagram and to give it priority in future searches.
Figure 4 shows the graphical user interface that presents the results obtained after executing the query according to the parameters entered.

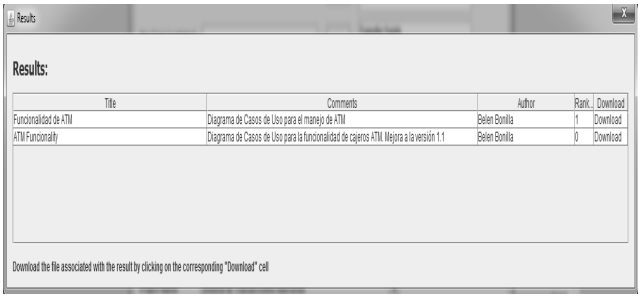



Fig. 4 GUI for Results Presentation

After the XMI file that defines the use case diagram required is downloaded, it can be opened through any UML modeling tool.
For example, Figure 5 shows how an XMI file that was recovered by the tool can be imported and displayed in Rational Rose, ArgoUML and StarUML.

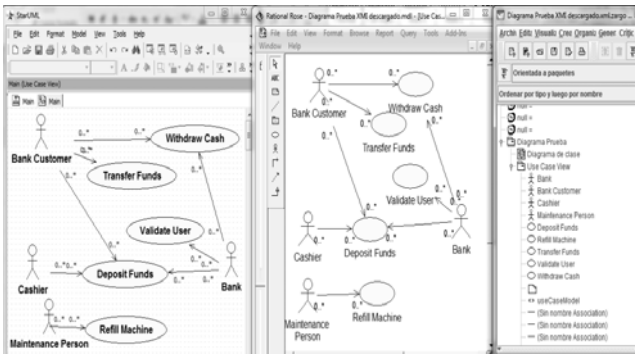



Fig. 5 Visualization of an XMI File retrieved through the tool

## 4. Conclusions

In this paper we present an approach to reuse use case diagrams by storing their information in OWL ontology and the implementation of a tool in Java, using Semantic Web technologies and tools like Jena framework and SPARQL query language. This tool allows querying individuals of the OWL ontology and retrieving associated use case diagrams, in XMI format, according to the users' input parameters through the GUI search.
The storage of the information of use cases diagrams in an ontology allows adding semantic to their definition, which benefits the process of searching and retrieving them, as they leverage the capabilities of inference that have the OWL ontologies, feature that allows to generate knowledge from previous knowledge.
We believe that the main advantages obtained with the use of this tool are: saving time and effort during the stage of requirements analysis and modeling, and reliability in the use case diagrams that have been recovered.
It is very important to remark that this work is limited to the reuse of use case diagrams. Currently it does not work for other UML diagrams.
As future work, we plan to carry out a series of case studies that test the features and benefits of the tool. Then, we will work on extending the tool to work with any UML diagram, not only use case diagrams.

### Acknowledgments

Work funded by the Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT) of Panamá through the Proposal No. APY-GC10-024B.

## References

[1] C.W. Krueger, "Software Reuse". ACM Computing Surveys, Vol. 24, No.2, 1992, pp.131-183.
[2] W.N. Robinson, and H.G. Woo, "Finding Reusable UML Sequence Diagrams Automatically", IEEE Software, Vol.21, No.5, 2004, pp.60-67.
[3] Y. Kim, and E.A. Stohr, "Software Reuse: Survey and Research Directions". Journal of Management Information Systems, Vol.14, No.4, 1998, pp. 113-147.
[4] H. Eichelberger, "Automatic layout of UML use case diagrams", SoftVis '08 Proceedings of the 4th ACM symposium on Software visualization. Munich, Germany, 2008, pp.105-114.
[5] A. Cockburn, Writing Effective Use Cases. Addison-Wesley Longman Publishing Co. Inc., Boston, MA. 2000.
[6] M.D. Monegan, An Object-Oriented Software Reuse Tool. Technical Report, Massachusetts Institute of Technology Cambridge, MA, USA, 1989.
[7] G. Gicca, "Reuse system software repository tool concepts", ACM SIGAda Ada Letters, Vol.11, No.1, 1991, pp. 70-89
[8] S. Henninger, "An Evolutionary Approach to Constructing Effective Software Reuse Repositories", ACM Transactions

on Software Engineering and Methodology (TOSEM), Vol.6, No.2, 1997, pp.111-140
[9] R.J. Chassell, An Introduction to Programming in Emacs Lisp. GNU Press - Free Software Foundation, Boston, MA, 2009
[10] M.C. Blok, and J.L. Cybulski, "Reusing UML Specifications in a Constrained Application Domain", APSEC 98 Proceedings of the Fifth Asia Pacific Software Engineering Conference. Washington, USA, 1998, pp. 196.
[11] W.N. Robinson, and H.G. Woo, "Finding Reusable UML Sequence Diagrams Automatically", IEEE Software, Vol.21, No.5, 2004, pp.60-67.
[12] D.J. Cook, et al., "Subdue: compression-based frequent pattern discovery in graph data", Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations (OSDM 05). Chicago, USA, 2005, pp.71-76.
[13] H. Happel, et al., "KOntoR: An Ontology-enabled Approach to Software Reuse", Proceedings of the Eighteenth International Conference on Software Engineering Knowledge Engineering (SEKE'2006). San Francisco, USA, 2006, pp. 349-354.
[14] OMG, 1997. UML Resource Page. [online] Disponible en: http://www.uml.org/
[15] I. Jacobson, et al., Object-Oriented Software Engineering – A Use Case Driven Approach, Addison – Wesley, 1992.
[16] OMG, 2005. Catalog of OMG Modeling And Metadata Specifications. [online] Disponible en: http://www.omg.org/cgi-bin/doc?formal/07-12-02
[17] Bechhofer, S. et al., 2004. OWL Web Ontology Language Reference. [online] Disponible en: http://www.w3.org/TR/owl-ref/
[18] G. Guizzardi, "The role of foundational ontologies for conceptual modeling and domain ontology representation", 7th International Baltic Conference on Databases and Information Systems. Vilnius, Lithuania, 2006, pp.17-25.
[19] W3C, 2005. Guía Breve de Web Semántica. [online] Disponible en: http://www.w3c.es/divulgacion/guiasbreves/websemantica
[20] JDOM Project, 2000. JDOM. [online] Disponible en: http://www.jdom.org/
[21] SourceForge, 2000. Jena - A Semantic Web Framework for Java. [online] Disponible en: http://jena.sourceforge.net/
[22] G. Lausen, et al., "Foundations of SPARQL query optimization", Proceedings of the 13th International Conference on Database Theory (ICDT '10). Lausanne, Switzerland, 2010, pp. 4-33.
[23] StarUML. StarUML - The Open Source UML/MDA Platform. [online] Disponible en: http://staruml.sourceforge.net/en/
[24] IBM, Rational Rose Enterprise, [online] Disponible en: http://www-01.ibm.com/software/awdtools/developer/rose/enterprise/index.html
[25] CollabNet, 2009, ArgoUML. [online] Disponible en: http://argouml.tigris.org/

**Belén Bonilla-Morales** is a student of the Master of Science in Information and Communication Technology at the Technological University of Panama. She was awarded a Bachelor's degree in Engineering and Computer Science from the Technological University of Panama in 2009. Her research interests include Software Engineering, Semantic Web, Grid Computing and other topics.

**Sérgio Crespo** is a professor at the Universidade do Vale do Rio Dos Sinos - Brasil, PhD awarded by Pontifícia Universidade Católica do Rio de Janeiro.

**Clifton Clunie** is a professor at the Technological University of Panama, PhD awarded by the Universidade Federal do Rio de Janeiro.

# 3D Media over Future Internet: Current Status and Future Research Directions

**Tasos Dagiuklas**[1]
**[1] CONES Research Group, Dept. of Telecommunication Systems and Networks, TEI of Mesolonghi Nafpaktos, 30300, Greece**

**Abstract**
Future Media Internet has been designed to overcome current limitations and address emerging trends including: network architecture, content and service delivery across heterogeneous networks, diffusion of heterogeneous nodes and devices, mass digitisation, new forms of (3D) user centric/user generated content provisioning, emergence of software as a service and interaction with improved security, trustworthiness and privacy. This paper presents current and future research trends for 3D Video Delivery across the entire networked-media ecosystem (from the encoding/packetisation, through the transmission and up to end-user experience).
***Keywords:*** *3D Media Future Internet*

## 1. Introduction

The Internet was designed for purposes that bear little resemblance to today's usage scenarios and related traffic patterns. In the longer term, the exponential increase of the user-generated multimedia content and the number of mobile users will raise many new challenges. In this respect, Future Media Internet will not simply be a faster way to go online [1]. It will be designed to overcome current limitations and to address emerging trends including: network architecture, content and service mobility, diffusion of heterogeneous nodes and devices, mass digitisation, new forms of (3D) user centric/user generated content provisioning, emergence of software as a service and interaction with improved security, trustworthiness and privacy.

There are a lot of advances in 3D media technologies in terms of capturing, representation, coding, delivering, and visualization. 3D media has been evolving in various areas, covering various market segments such as professional (e.g. scientific, medicine, education, training etc) and entertainment (3D Gaming, 3D Live Sports, 3D Live Shows etc) sectors. These new technologies give the ability to design and develop new types of applications ranging from Virtual Collaborative Environments (e.g. multi-modal interactions, etc) to Edutainment (e.g, the user is allowed to choose a service in terms of content production, searching program content, provide service according to user's preferences, tailor education content according to user's need).

The successful deployment of 3D Media is based on issues related to efficient transport of content, as well as, effective and inexpensive 3-D displays. However, although individual technologies are evolving fast, their integration and effective use is still limited by numerous open issues. In fact, while the production of 3D films, movies and live broadcasting events started in the early 50s, the research community has only recently paid attention at the transmission/broadcasting of 3D (rather than conventional 2D) applications and services. In the same time, latest advances on broadband network access (e.g. wireless: 3G, 4G, LTE, WiMax etc) and core (optical) technologies and digital broadcasting (e.g. DVBx) in terms of capacity, speed, ubiquity and reliability necessitates the optimization of the delivery of the aforementioned new 3D applications. These applications pose unique problems in comparison to natural 2D media coding data.

The aim of this paper is to outline the challenges associated with 3D Media Delivery across heterogeneous networks. This includes media functionalities (in terms of coding efficiency/scalability, error resilience, packetisation schemes, streaming protocols, adaptation, rate control, etc), network architectures to deliver 3D Video, use of rate adaptation in order to maximize QoS under network/user constraints with guarantee 3D viewing, wireless 3D Video Delivery and se of and error concealment for 3D Video

The paper is structured as follows: Section II present the current status in terms 3D Video Coding Standards, Transport Protocols and Wireless Media Delivery, Section II outlines future research challenges, issues and directions, Section IV presents the conclusions.

## 2. Current Status in 3D Video Communications

### 2.1 3D Video Coding

Research on video coding techniques for 3D is an increasing research topic around the world. A few coding techniques exist, and the most important ones are: 2D+Depth, as specified by ISO/IEC 23002-3 (and also referred to as MPEG-C Part 3) [2], and Multiview Video Coding (MVC), as specified by ISO/IEC 14496-10 | ITU-T Recommendation H.264 [3]. 2D+ Depth supports the inclusion of depth for generation of an increased number of views. While it has the advantage of being backward compatible with legacy devices and is agnostic of coding formats, it is only capable of rendering a limited depth range since it does not directly handle occlusions. Depth information can be also included as a layer in Scalable Video Coding (SVC).

Multiview Video Coding (MVC) supports the direct coding of multiple views and exploits inter-camera redundancy to reduce the bit rate. MVC gives very good 3D rendering capability, but the bit-rate of MVC encoded video is proportional to the number of views [4]. Powerful algorithms and open international standards for MVC and coding of video plus depth data are available and under development, which will provide the basis for introduction of various 3DTV systems and services in the near future [5]. The research area is relatively young when compared to 2D video coding. Therefore there is a lot of room for improvement and development of new algorithms and coding methods.

1. Enabling stereo devices to cope with varying display types and sizes, and different viewing preferences. This includes the ability to vary the baseline distance for stereo video to adjust the depth perception, which could help to avoid fatigue and other viewing discomforts.
2. MPEG also envisions that high-quality auto-stereoscopic displays will enter the consumer market in the next few years. Since it is difficult to directly provide all the necessary views due to production and transmission constraints, a new format is needed to enable the generation of many high-quality views from a limited amount of input data, e.g. stereo and depth. The 3DV format is expected to have several advantages in terms of bit rate and 3D rendering capabilities.
   - 2D+Depth, as specified by ISO/IEC 23002-3, supports the inclusion of depth for increasing the number of view. It exhibits backward compatibility with legacy devices and is agnostic of coding formats. It is only capable of rendering a limited depth range since it does not directly handle occlusions. The 3DV format expects to enhance the 3D rendering capabilities beyond this format.
   - Multiview Video Coding (MVC), as specified by ISO/IEC 14496-10. ITU-T Recommendation H.264, supports the direct coding of multiple views and exploits inter-camera redundancy to reduce the bit rate. Although MVC is more efficient than simulcast, the rate of MVC encoded video is proportional to the number of views. The 3DV format expects to significantly reduce the bit rate needed to generate the required views at the receiver.
   - Multiview plus depth (MVD): It regards rendering techniques at the receiver providing great adaptation to varied depth experienced from different 3D Displays. This involves view synthesis by interpolating color information from multiple views. MVD can be extended to cope with large changes in view conditions by sending multiple MVD streams [6].
   - Layered depth Video (LDV): It is proposed in order to further reduce the bit rate from MVD. It projects the central camera view into other neighbouring views and then determine the difference between projection and neighbouring camera result as residual information [7], [8].

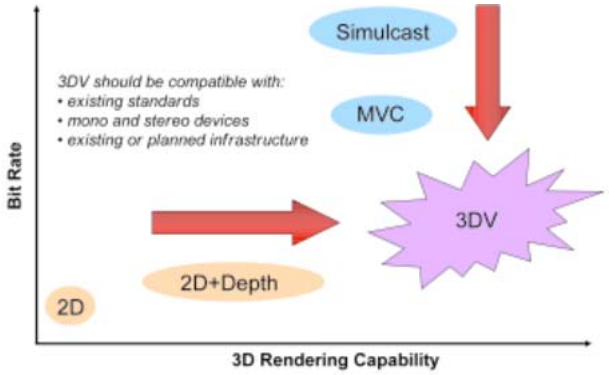



**Figure 1: Towards 3DV Evolution (Source : http://multimediacommunication.blogspot.com)**

### 2.2 Transport Protocols for 3D Video

IETF has adopted MVC as the 3D video coding standard to convey 3D video packets. MVC exhibits packetization similarities with H.264 SVC [9]. There are three transmission modes for multiple-view video, referring to Single Session Transmission (SST), Multi-Session Transmission (MST) and Media-Aware Network Element (MANE)-based transmission [10]. In SST, all

MVC packets are carried in a single RTP session utilizing only a single transport address/port. MANE generally resides in the path between the server and the clients. The server keeps using MST transmission. However in this case, MANE collects all RTP sessions and de-packetize them. It then customizes NAL Units according to the client's needs through adaptation decision taking engine (ADTE) and aggregates new packets for SST to clients through a single transport address using in a single RTP session. MANE is able to take important information using either signaling or RTP and NAL Unit Headers, as illustrated in the following figure.

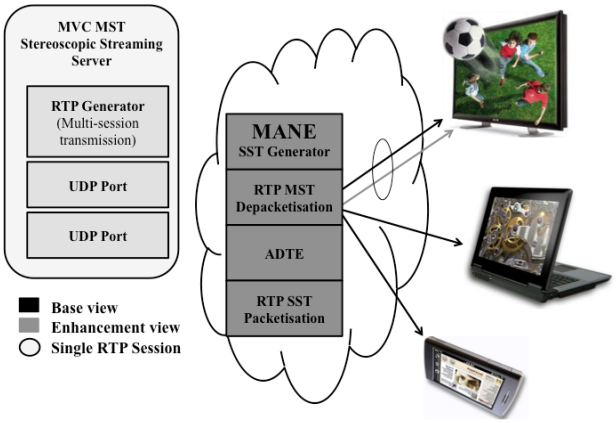



**Figure 2: MVC Packetization Modes for 3D Video**

The state-of-the-art video streaming protocols are RTP/UDP/IP, while the next generation protocol is expected to be RTP/DCCP/IP [11]. In particular, RTP/UDP does not contain any congestion control mechanism and, therefore, can lead to congestion collapse when large volumes of multi-view video are delivered. On the other hand, the datagram congestion control protocol (DCCP) is designed as a replacement for UDP for media delivery, running directly over the IP to provide congestion control without reliability [12]. DCCP is a transport protocol that implements bi-directional unicast connections of congestion-controlled, unreliable datagrams. DCCP provides reliable handshakes for connection setup/tear down and reliable negotiation of options. Besides handshakes and feature negotiation, DCCP also accommodates a choice of modular congestion control mechanisms. There exist two congestion control schemes defined in DCCP currently, one of which is to be selected at connection start-up time. These are TCP-like Congestion Control [13] and TCP-Friendly Rate Control (TFRC) [14]. TCP-like Congestion Control, identified by Congestion Control Identifier 2 (CCID2) in DCCP, behaves similar to TCP' s Additive Increase Multiplicative Decrease (AIMD) congestion control, halving the congestion window in response to a packet drop. Applications using this congestion control mechanism will respond quickly to changes in available bandwidth, but must tolerate the abrupt changes in the congestion window size typical of TCP. On the other hand, TFRC, which is identified by CCID3, is a form of equation-based flow control that minimizes abrupt changes in the sending rate while maintaining longer-term fairness with TCP. It is hence appropriate for applications that would prefer a rather smooth sending-rate, including streaming media applications with a small or moderate receiver buffer. In its operation, CCID3/TFRC calculates an allowed sending rate, called the TFRC rate, by using the TCP throughput equation, which is provided to the sender application upon request. The sender may use this rate information to adjust its transmission rate in order to get better results.

2.3 3D Video Networking Architectures

Transmission of video over the Internet is currently an active research and development area, where significant results have already been achieved. There are already video-on-demand services, both for news and entertainment applications, offered over the Internet. Also, 2.5G and 3G mobile network operators started to use IP successfully to offer wireless video services. Looking at these advances, the transport of 3DTV signals over IP packet networks seems to be a natural choice. The IP itself leaves many aspects of the transmission to be defined by other layers of the protocol stack and, thus, offers flexibility in designing the optimal communications system for various 3D data representations and encoding schemes. 3DTV streaming architectures can be classified as:

1. server unicasting to a single client:
2. server multicasting to several clients.
3. P2P multicasting, where each peer forwards packets to several other peers: Peer-to-peer (P2P) technology has the potential to provide a more cost effective and flexible delivery solution for future 3D entertainment services [15]. In terms of architecture definitions, two main architectures have been used, namely:
   - Mesh-based approaches: They derive from file sharing applications, peers organize themselves into a mesh, independently requesting pieces, or chunks, of the video content from neighbors, without any regard for the structure of the distribution path
   - Tree-based approaches: In tree-based approaches, video packets are forwarded along a pre-determined path forming multicast trees. The advantage of combining more than one multicast trees is that the robustness of the system is increased, due to path diversity.

## 2.4 Mobile 3D Media Delivery

The term 'handover' (HO) refers to the association of a Mobile Node (MN) from an old Point of Access (PoA) towards a new PoA wither belonging to the same access network or to another one. The mobility concept is defined in different ways by the relevant standardization fora [16]:

1. Nomadism: It is the ability of the user to change his network point of attachment while the end-user is on the move. When the network point of attachment is changed, the user's service session is completely stopped and it can be resumed later on.
2. Session Continuity: It refers to the ability that the end-user's terminal can switch to a new network point of attachment while maintaining the on-going session from the old point of attachment to the new one. This may include a session break and resume, or a certain degree of service interruption or loss of data while changing to the new access point."
3. Seamless handoff: the handoff algorithm should minimize the packet loss. It is sometimes referred to as smooth handoff. Transparent migration of on-going data flows between two access points belonging to independent heterogeneous technologies is achievable, and tools and mechanisms for supporting this type of mobility should be placed within the next generation networking architectures.

The co-existence of multiple wireless mobile clients that can support different networking technologies and multimedia services with heterogeneous requirements is becoming a common trend in future wireless era [17]. However, these wireless networks bear a higher level of randomness than their wired counterparts. Handling Mobility in an IP environment can be supported by Mobile IP and its extensions (Hierarchical Mobile IP, Mobile IPv6 and Fast Mobile IP with Handoffs) [18]-[21]. However, in a mobility environment comprising heterogeneous networks the support and maintenance of on-going sessions with strict QoS requirements may be a challenging task.

This necessitates the use of frameworks where handover can be either by the MT or the network through a cooperative synergy. This synergy may require the capturing of information/statistics from Physical, Network and Application Layer. This concept has been introduced under the Media Independent Handover Framework within IEEE 802.21 [22], [23]. When a node/client detects a deterioration of the received signal, it may initiate a procedure for registering itself on a different RAT, based on the available infrastructure. Such a Handover decision can be triggered by functions such as the Media Independent Handover (MIH) entity which decides on where to direct the client based on information coming from network discovery (e.g. best candidate AP, SNR), or based on information from network and the application server (e.g. QoE). Next, it applies this information to a pre-defined handover algorithm in order to decide which AP is the optimal solution for the handover [24].

Handover across heterogeneous RATs (Radio Access Technologies networks) of mobile devices with on-going 3D Video sessions is a challenging task. This is due to the fact of the number of views required by the end-user and the bit rate requirements for each view. Therefore in order to support and maintain QoS of a going-session the seamless mobility necessitates the transfer of each view towards a new wireless link. Due to the bandwidth constraints that are imposed by the different wireless networks different mobility strategies options will be considered:

1. Exploit mobility with path diversity so that each view may be rerouted to a different radio access network, combine handover with rate adaptation/transcoding in case of network bandwidth limitations.
2. Priority handling will be given to the base view (selecting the network that can accommodate the base view rate without deteriorating QoS and violating fairness of other connections).

The following figure illustrates an architecture how to support seamless mobility for 3D Video.

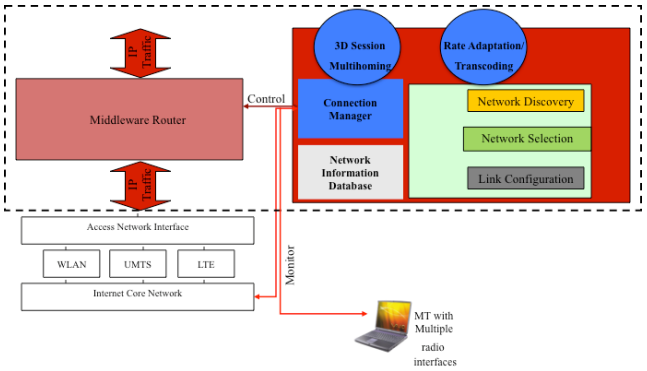



**Figure 3: Seamless Mobility Architecture for 3D Video**

## 2.3 Rate Control for 3D Video

Rate adaptation of multiview video may be achieved at constant perceived 3D video quality by adaptation of the spatial, temporal and/or Signal-to-Noise (SNR) resolution of one of the views while encoding and transmitting the other view at full rate. Several open loop and closed loop rate adaptation strategies for stereo and multiview video at the server and the client side are studied for UDP and DCCP protocols. In the closed loop rate adaptation, each client estimates some function of the

received signal and feeds it back to the transmitter. The transmitter determines an optimized rate for the next transmission based on the received feedback. On the other hand, in the open loop rate adaptation, the transmitter does not use any feedback from the receiver.

Adequate perceptual metrics are necessary for 3D video quality in adaptation scenarios where different types of underlying communication channels might be used [25], [26]. The temporal effect of errors in depth perception is not yet fully known which leaves open questions about the subjective impact of freezing depth perception at the receiver during limited adaptation periods. The behavior of a 3D video system and the user perception is different when switching either to still frame or monoscopic view in case one stream is lost [27], [28]. An adaptation engine based on the subjective impact of such behavior would certainly yield better perceptual quality.

### 2.4 Error Concealment for 3D Video

3D Video Artifacts can be classified into the following categories: structure, color, motion and binocular [30]. These artifacts such as blurriness, distortion, aliasing, geometry distortion and cross-talk affect human perception on image structure such as contours and textures. The errors can be classified as error in shape and error in appearance [31].

Streaming media applications often suffer from packet losses at both wireline and wireless networks. Such losses may be due either to congestion or physical layer impairments. Error Concealment can be used in order to recover from errors and improve the perceived video quality. Such mechanisms may be grouped at the following categories:

1. Using Error Resilience mechanisms at the encoder part. There are several approaches how error resilience can be applied at the 3D Video that depend on 3D technology being used:
   - In MVC, it is exploited the fact that certain number of views are needed to each user at any time [11], [32], [33], [34], [35].
   - In 2D+V, 3D Video can be encoded with 10%-20% extra bit rate by exploiting the correlation between the video and the depth information [5], [6]
   - Use MDC, where the video is transmitted as several independent descriptors. Each descriptor may correspond to each view [36], [37]. Upon the reception of only one description, the bit stream can be decoded to construct a lower-quality representation.
2. Forward Error Correction: Using Unequal Error Protection (UEP), different portions of the 3D video bit stream are protected, depending on the relative importance of the views [38]. Flexible Marcoblock Ordering (FMO) could be used in order to classify the importance of bits that are generated by the encoder so that Unequal Error Protection Techniques can be applied to the encoder.
3. Exploit Temporal Correlations in 3D Domain: For stereo videos, temporal correlations within each view, in addition to inter-view correlations, can be used to conceal lost blocks or frames more effectively. [39], [40], and [41] propose error concealment algorithms for stereo videos can be employed to conceal erroneous regions in multi-view video sequences

## 3. Future Research Challenges

### 3.1MVC Coding

In the area of MVC coding, future enhancements regards MVC extensions to achieve highly flexible and scalable compressed representation. The overhead scalability should be kept to as low as possible as compared to non-scalable compressed 3D multi-view video. Similar to 2D SVC, scalable coding can offer flexibility in terms of scheduling to diverse networks. Each description will be scalable so that all descriptions can be efficiently adapted to the available rate of each link for effective congestion control. This necessitates the adoption of a framework for selecting the best encoding configuration for MD-SMVD, which will strike the best balance between minimizing the average end-to-end rate-distortion performance of each description given a set of packet loss probabilities and minimizing overall redundancy and maximizing the range of extraction points of each scalable description. The optimization variables will be some SMVD encoding parameters (e.g. layer QP values, macroblock modes, inter-layer prediction modes) and MD generation alternatives that result in different levels of redundancy at a fixed total rate for all descriptions

### 3.2 Novel 3D Video Networking

Network Coding is a promising technique that could be used for network content distribution [42], [43]. The concept behind network coding relies on the following remark. Communication networks today share the same fundamental principle of operation: whether it is packets over the Internet or signals in a phone network,

information is transported in the same way as cars share a highway or fluids share pipes. That is, independent data streams may share network resources, but the information itself is separate. Routing, data storage, error control, and generally all network functions are based on this assumption. Network coding breaks this assumption. Instead of simply forwarding data, nodes may recombine several input packets into one or several output packets. It can be proven that the theoretical throughput within the network by applying coding (linear combinations of different pieces from the original content). In a large distributed cooperative system finding an optimal packet propagation scheme that minimizes the client download time is very difficult. This is especially the case in practical systems that cannot rely on a central scheduler and, instead, allow nodes to make local decisions. The scheduling problem becomes increasingly difficult as the number of nodes increases. A key question is the selection of network code (XOR operation among the packets) in order to optimize throughput [44], [45]. One important constraint with network coding is that as bandwidth efficiency increases, longer delays may be applied to some packets due to longer queuing time of packets, to allow packet losses occurrence. There is a lot of interest within the research community how network coding could be used in multimedia communications. This necessitates considering coding distortion conveyed in a video packet o construct the network information flows [46], [47]. Given that there are K downstream nodes with different packet loss rates, a key question is to design optimal scheduling algorithms that determine which packets (pure and mixed packets) are transmitted at certain times in order to maximize video quality. The key question that is raised regards the network code (XOR operation among the packets) so that both video distortion and throughput is optimised. The distortion of each packet can be determined by the source and the view type and communicated to the distribution nodes in order to get transmission in a rate-distortion optimised manner [47].

The employment of network coding paradigm will provide efficiency in terms of throughput within the core/distribution network, distributing large chunk of MVC 3D video packets across different edges where heterogeneous wireless networks (e.g. 3G, WLAN, WiMAX and LTE) are located. It is important to consider network coding algorithms that are video aware in terms of the importance of the packets that would significant improve throughput. It is important to consider how network coding could be used in an efficient manner for Scalable MVC. The advantage of scalable video coding necessitates the use of strategies that combine packets of the same important to the network information flow. This necessitates exploiting video packets levels of importance (e.g. in MVC packets from base view are important in order to reconstruct non-base view) and how this information can be correlated with network codes so that several packets from different views can be merged together increasing throughput. Two scenarios can be exploited. In the first one, packets from one view can construct a network flow, whereas in the second case packets from the same layer from multiple views can construct an information flows. These ideas are presented in the following figure, by considering two layers and two views.

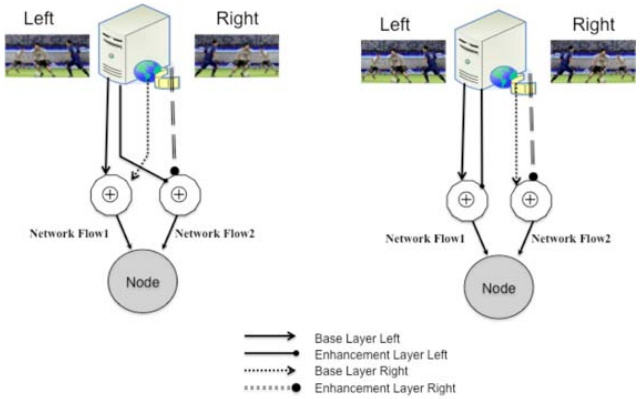



**Figure 4: Network-Aware Network Coding for Scalable 3D MVC Video**

## 3.3 Error Concealment for 3D Video

Error Concealment Schemes for 3D Video could fall in the following categories:

1. Network Assisted Mechanisms where out-band signalling protocols may be used in order to send parameter sets and important information (i.e. GOP size, Number of Views, NAL Unit Headers, View id) that are important for the reconstruction of the 3D View.
2. Concealment methods for missing depth information due to either transmission errors or packet loss. Existing spatial and temporal interpolation techniques will be subject to further research and development in order to achieve efficient perceptual concealment of depth artifacts in 3D video. The additional level of dependency between views also brings a new dimension into the problem of error propagation, which must be taken into account.
3. Methods for restoration of compressed depth maps affected by blocking artifacts and reconstructed missing areas. The 3D perceptual quality will be evaluated using both objective metrics and subjective testing.

4. Algorithms for rendering occluded areas and inpainting will be investigated, in order to provide reconstructed views with a higher fidelity.
5. UEP mechanisms based on dynamic combination of FEC and depth information will be devised to increase the quality of 3D video delivered over error prone channels. Compressed depth maps along with light forward error correction are transmitted for increased error resilience and improved error concealment at the used terminal.
6. Error concealment methods based on structural and motion correlations of 2D plus depth video will be investigated. Moreover, the redundant information available in adjacent views will be employed to restore missing image blocks in both 2D and depth map video.

### 3.4 Quality of Experience

QoE is inherently a subjective concept. The ultimate way of measuring QoE is to obtain subjective measurements of the different dimensions of QoE. Subjective evaluations attempt to assess the perceived quality of audio and video multimodal networked services. Several methodologies for subjective evaluation of video quality, sound quality and multimedia quality for particular applications have been standardized within the ITU, e.g. recommendation ITU-R BT.500-11 [48], which deals with assessment of television picture quality, and ITU-T P.910 [49], which deals with video quality in multimedia applications. Moreover, ITU-T Recommendation P.911 describes subjective assessment methods for evaluation of audiovisual quality in multimedia applications, P.920 includes aspects relevant for interactive applications [51] and ITU-R Recommendation BS-1284 specifies general methods for the subjective assessment of sound quality.

On the other hand, objective measurement methods analyse and measure characteristics of the source and/or the reconstructed processed media signal by predicting the perceived quality of audio, video and speech. An extensive coverage of existing models for predicting video quality in both a double-ended and single-ended fashion is provided by Winkler, in his book, entitled "Digital Video Quality – Vision Models and Metrics", published by John Wiley in 2005.

Visual attention utilises a bottom-up control model of visual attention in primates, as introduced in [52] and [53]. This can be accomplished by a decomposition of the input image into a set of multi-scale neural "feature maps" which extract local spatial discontinuities in the modalities of colour, intensity and orientation. Compared to the underlying original psychological attention model of Treisman [54], the feature "stereo-distance" is still missing in [55]. Additionally, Treisman distinguishes between top-down and bottom-up features. Top-down features lead the search for salient objects by prior knowledge on context and / or object properties.

Attention Models using top-down information are rarely available in literature and mainly known from robotics. Novel visual attention models could be used to detect important regions by extracting salient image features. These regions of interest could be used for region based coding and guidance of the content adaption process.

3D immersive environments pose the unique situation of service operators having to deliver multiple multimedia types directly to the user. Interfacing with the users avatar within the environment would allow the user full control of their QoE and make the architecture truly adaptive. However this is not the only way QoE could be attained in such environments. Certain activities, which take place within the environment, can require different levels of QoE. For instance when an user's avatar is interfacing with the 3D environment itself would require a different level of QoE in certain medias than if a user's avatar was interacting with another user. Thus, due to the subjective quality of the interaction itself users can be satisfied in di□erent ways. The wide range of applications used today on mobile devices means that many different levels of service are required to satisfy users.

## 4. Conclusions

This paper surveys the current and future research trends associated with Next generation 3D Media Applications over Future Internet. Particular emphasis has been given on the media encoding and delivery of 3D Applications across heterogeneous content-aware networks (e.g. wireless such as LTE, WiMAX, DVB and wireline such as xDSL, FTTx), in an ecosystem across the entire value chain of Networked Media (from the encoding/packetisation, through the transmission and up to end-user).

## References

[1] P. Daras, F. Alvarez,: A future perspective on the 3D Media Internet. In: Tselentis, G.e.a. (ed.) Towards the Future Internet - A European Perspective, pp. 303–312. IOS Press, Amsterdam (2009)
[2] ISO/IEC JTC1/SC29/WG11, "ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information", Doc. N8259, Klagenfurt, Austria, July 2007.

[3] ISO/IEC 14496-10:2005/FDAM 3 Scalable Video Coding, Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, Lausanne, N9197, Sep. 2007.
[4] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The Emerging MVC Standard for 3D Video Services", EURASIP Journal on Advances in Signal Processing, Volume 2009
[5] A. Smolic, et al, "Coding Algorithms for 3DTV: A Survey", IEEE Trans. Ob Circuits and Systems for Video Technology, Vol. 17, November 2007
[6] P. Kauff et al,"Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability", Signal Processing: Image Communication, Vol. 22, February 2009
[7] K. Muller et al, ""View synthesis for advanced 3D video systems", EURASIP Journal on Image and Video Processing, 2008
[8] K. Muller et al, "3D Visual content compression for communications", IEEE Multimedia Communications E-Letter, Vol. 4, July 2009
[9] IETF Draft, "RTP Payload Format for SVC Video", draft-ietf-avt-rtp-svc, October 2010
[10] K.-D. Seo, J.-S. Kim, S.-H. Jung, and J.-J. Yoo, "A Practical RTP Packetization Scheme for SVC Video Transport over IP Networks", ETRI Journal, No. 2, April 2010, DOI: 10.4218/etrij.10.1409.0031
[11] A. M. Tekalp, E. Kurutepe and M. R. Civanlar, "3DTV over IP", IEEE Signal Processing, November 2007
[12] IETF RFC 4340, "Datagram Congestion Control Protocol", March 2006
[13] IETF RFC 4341, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control", March 2006
[14] IETF RFC 4342, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)", March 2006
[15] Y. Liu, Y. Guo and C. Liang, "A survey on peer-to-peer video streaming systems", *Peer-To-Peer Networking and Applications*, Vol. 1, 2008
[16] N. Passas et al, "Architectures and protocols for mobility management in All-IP Mobile Networks", IEEE Wireless Magazine, April 2008
[17] Y. M. Fang, "Special Issue: Seamless Content Delivery in the Future Mobile Internet", IEEE Wireless Communications Magazine, October 2009
[18] IETF RFC, "IP Mobility for IPv4", March 2002
[19] IETF RFC 4140, "Hierarchical Mobile IPv6 Mobility Management", August 2005
[20] IETF RFC 4068, "Fast Handovers for Mobile IP", July 2005
[21] IETF RFC 4721, " Mobile IPv4 Challenge/Response Extensions", January 2007
[22] IEEE 802.21/D10.0. Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services (Work in progress). Technical report, IEEE Draft, April 2008
[23] G. Lampropoulos, A.K. Salkintzis, and N. Passas, "Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks," IEEE Communications Magazine, Vol. 46, January 2008.
[24] J. Rodriguez, M. Tsagkaropoulos, I. Politis, T. Dagiuklas and S. Kotsopoulos, "A Middleware Architecture Supporting Seamless and Secure Multimedia Services across Inter-Technology Radio Access Network", IEEE Wireless Communications Magazine, October 2009
[25] G. B. Akar, M. Tekalp, C. Fenn and R. Civanlar, "Transport methods in 3DTV- a survey", IEEE Trans. On Circuits and Systems on Video Technology, Vol. 17, November 2007.
[26] A. Aksay et al, "End-to-end stereoscopic video streaming system with content-adaptive rate and format control", Signal Processing: Image Communication, Vol. 22, February 2007
[27] S. Jolly, M Armstrong and R Salmon, The Challenges of Three-Dimensional Television, White Paper WHP 173, BBC, January 2009
[28] O. Stankiewicz, K. Wegner, M. Domański, "Error Concealment for MVC and 3D Video Coding", Picture Coding Symposium 2010,Nagoya Japan, December 2010
[29] A. Boev, D. Hollossi, A. Gotchev and K. Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content", Proc. Of SPIE, January 2009
[30] A. Boev, D. Hollossi, A. Gotchev and K. Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content", Proc. Of SPIE, January 2009
[31] J. Kilner, J. Starck, J. Y. Guillemaut and A. Hilton, "Objective quality assessment in free-viewpoint video production", Signal Processing: Image Communication, Vol. 24, January 2009
[32] G. B. Akar, M. Tekalp, C. Fenn and R. Civanlar, "Transport methods in 3DTV- a survey", IEEE Trans. On Circuits and Systems on Video Technology, Vol. 17, November 2007.
[33] X. Cao, Y. Liu, Q. Dai, "A flexible client-driven 3DTV system for real-time acquisition, transmission and display of dynamic scenes", EURASIP Journal on Advances in Signal Processing, Vol. 2009
[34] E. Kurutepe, and T. Sikora, "Feasibility of multi-view video streaming over P2P networks", IEEE 3DTV Conference, Instabul, Turkey, May 2008
[35] J. Lou, H. Cai and J. Li, "Interactive multiview video delivery based on IP Multicast", EURASIP Advances in Multimedia, Volume 2007
[36] A. Norkin et al, "Schemes for multiple description coding of stereoscopic video", LNCS on Multimedia Content Representation, Classificatoon and Security 2006
[37] H. A. Karim, 'Scalable Multiple Description Video Coding for Stereoscopic 3D', IEEE Transc. On Consumer Electronics, May 2008
[38] A. Tan, A. Aksay, G. Akar and E. Arikan, "Rate-distorton optimization for stereoscopic for 3D transmission", EURASIP Journal on Advances in Signal Processing, Vol. 2009
[39] K.A.M. Guenther, C. Clemens, T. Sikora, A fast displacement-estimation based approach for stereoscopic error concealment, Proc. PCS 2004, Oregon, USA, November 2004
[40] L. Pang, M. Yu, W. Yi, G. Jiang, W. Liu, Z. Jiang, Relativity analysis-based error concealment algorithm for

entire frame loss of stereo video, in: Proc. International Conference on Signal Processing, 2006.
[41] X. Xiang, D. Zhao, Q. Wang, X. Ji, W. Gao, A novel error concealment method for stereoscopic video coding, in: Proc. ICIP'07, 2007, pp. 101–104.
[42] R. Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung, "Network Information Flow", IEEE Trans. On Information Theory, Vol. 46, July 2000
[43] H. Wang, J. Liang and C.-C. J. Kuo, "Overview of Robust Video Streaming with Network Coding", *Journal of Information Hiding and Multimedia Signal Processing*, January 2010
[44] S. Katti et al, "XORs in the air: Practical network coding", *ACM SIGCOMM*, Pisa, Italy, September 2006
[45] R. Koetler and M. Medard, 'An algebraic approach to network coding', IEEE/ACM Trans. On Networking, 2003
[46] IEEE COMSOC MMTC E-Letter, "Network Coding for Multimedia Communications", March 2010
[47] H. Seferoglu and A. Markopoulou, 'Video-Aware Opportunistic Network Coding over Wireless Networks", IEEE JSAC, Vol. 27, No.5, pp. 1-16
[48] Recommendation ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Geneva, 1974-1997
[49] Recommendation ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," ITU-T, Geneva, 1996.
[50] Recommendation ITU-T P.911, "Subjective audiovisual quality assessment methods for multimedia applications" ITU-T, Geneva, 1998.
[51] Recommendation ITU-T P.920, "Interactive test methods for audiovisual communications," ITU-T, Geneva,1996.
[52] S. Frintrop, "VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search", Lecture Notes in Artificial Intelligence, vol. 3899, 2006
[53] Yao Zhang, Chun Yuan, Yuzhuo Zhong: Implementing DRM over Peer-to-Peer Networks with Broadcast Encryption. In: Proceedings of the 8th Pacific Rim Conference on Multimedia, pp. 236-245, 2007.
[54] A. Treisman, "Features and Objects in Visual Processing," Scientific American, vol. 255, pp. 106 – 115, 1986.
[55] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, pp. 1254–1259, 1998

**Tasos Dagiuklas** received the Engineering Degree from the University of Patras-Greece in 1989, the M.Sc. from the University of Manchester-UK in 1991 and the Ph.D. from the University of Essex-UK in 1995, all in Electrical Engineering. Currently, he is employed as Assistant Professor at the Department of Telecommunications Systems and Networks, Technological Educational Institute (TEI) of Mesolonghi, Greece.
He is the Leader of the Converged Networks and Services Research Group. He is also Senior Research Associate within the Wireless Telecommunications Laboratory of the Electrical and Computer Engineering Department at the University of Patras, Greece. Past Positions include teaching Staff at the University of Aegean, Department of Information and Communications Systems Engineering, Greece, senior posts at INTRACOM and OTE, Greece. He has been involved in several EC R&D Research Projects under FP5, FP6 and FP7 research frameworks, in the fields of All-IP network and next generation services.
He has served as TPC member to more than 30 international conferences. His research interests include Future Internet architectures and converged multimedia services over fixed-mobile networks. Dr Dagiuklas has published more than 100 papers at international journals, conferences and standardisation fora in the above fields.

# Determining Semantically Equivalent Questions in a Vietnamese Language Based Document Retrieval System

**Dang Tuan NGUYEN, Trung TRAN**

**Faculty of Computer Science, University of Information Technology, Vietnam National University – HCMC**
**Ho Chi Minh City, Vietnam**

**Abstract**

Generally speaking, in the natural language based document retrieval systems, the ability to determine questions having the equivalent meanings is an important requirement, it allows the optimization of the processing mechanism of questions in related systems. To determine Vietnamese questions having the equivalent meanings for a Vietnamese language based document retrieval system, we have to solve the problems: what is the meaning of a question? And, how can we determine the equivalent meanings of these questions? In this research, we established the method which can allows to determine the Vietnamese questions having the equivalent meanings.

***Keywords:*** *Natural Language Processing, Document Retrieval, Syntax, Semantics.*

## 1. Introduction

Generally speaking, in the natural language based document retrieval systems, the ability to determine questions having the equivalent meanings is an important requirement, it allows the optimization of the processing mechanism of questions in related systems. This is also a major challenge for the documents retrieval systems based on Vietnamese Question – Answering model, such as the systems built in [2], [3], [4], [5], [6].

To determine Vietnamese questions having the equivalent meanings for a Vietnamese language based document retrieval system, we have to solve the problems: what is the meaning of a question? And, how can we determine the equivalent meanings of these questions? In this paper, based on rules proposed in "Syntactic Structures" theory of Noam Chomsky [1], we focus on building and testing a method which allows to determine Vietnamese questions having the equivalent meanings, in the domain of document retrieval field.

## 2. Method of Determining Vietnamese Questions having Equivalent Meanings

In this research, we recognize the hypothesis:

- Hypothesis 1: There are Vietnamese questions which their meanings are considered equivalent in the Vietnamese language based document retrieval system.
- Hypothesis 2: The determination of Vietnamese questions having the equivalent meanings have to be implemented based on proper linguistic theory.

Based on "Syntactic Structures" theory of N. Chomsky [1], we developed a method which determines questions having the equivalent meanings for Vietnamese language based document retrieval system. The method that we propose is presented as follows:

Step 1. Analyze the syntactic structures of Vietnamese questions.

Step 2. Determine the syntactic structure of kernel sentence of each question.

Step 3. If these questions are transformed from the same kernel sentence, then determine the sequences of transformations applied to the kernel sentence for generating the considering questions. If these sequences of transformations belong to one of the alternatives determined in Table 1, then the considering questions are considered that they have the equivalent meanings.

In the proposed method above, we recognize that the meaning of a question is regulated by its deep structure. The deep structure of a question is generated by the phrase structure part. In Vietnamese, we do not have morphophonemic rules, therefore the deep structure is also

the structure of kernel sentence which is transformed from the structure of terminal string.

Following Chomsky's opinion in "Syntactic Structures" [1], the transformations applied for a kernel sentence do not change the meaning of the sentence. However, in this research, we recognize the additional hypothesis: there are transformations considered equivalent when they are applied to kernel sentences having proper structures.

In Table 1, we present the sequences of equivalent transformations in the sense of if these sequences of equivalent transformations have been applied to the same kernel sentence having the determined syntactic structure, then the generated sentences are considered having the equivalent meanings.

Table 1: The sequences of equivalent transformations

| | *Structure of kernel sentence* | *Sequences of equivalent transformations* |
|---|---|---|
| Group 1 | danh_ngữ_một + động_từ + danh_ngữ_hai | Tq<br>Tq → Tpass |
| | | Tq → Tw danh ngữ một<br>Tq → Tw danh ngữ một<br>→ Tpass |
| | danh_ngữ_một + động_từ + danh_ngữ_hai + định_ngữ | Tq → Tw danh ngữ hai<br>Tq → Tw danh ngữ hai<br>→ Tpass |
| Group 2 | danh_ngữ_một + động_từ + danh_ngữ_hai + bổ_ngữ | Tq<br>Tq → Tpass |
| | | Tq<br>Tq → Tobsep → Tpass |
| | | Tq → Tw danh ngữ một<br>Tq → Tw danh ngữ một<br>→ Tpass |
| | | Tq → Tw danh ngữ một<br>Tq → Tw danh ngữ một<br>→ Tobsep → Tpass |
| | | Tq → Tw danh ngữ hai<br>Tq → Tw danh ngữ hai<br>→ Tpass |
| | | Tq → Tw danh ngữ hai<br>Tq → Tw danh ngữ hai<br>→ Tobsep → Tpass |
| Group 3 | danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng<br><br>danh_ngữ_một + định_ngữ + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng | Tq<br>Tq → Tnominalize |
| Group 4 | danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai<br><br>danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + định_ngữ | Tq → Tw danh ngữ hai<br>Tq → Tw danh ngữ hai<br>→ Tnominalize |
| | | Tq → Tw danh ngữ một<br>Tq → Tw danh ngữ một<br>→ Tnominalize |
| | | Tq → Tw danh ngữ hai<br>Tq → Tw danh ngữ hai<br>→ Tnominalize |

Transformations Tpass, Tq, Tw, Tobsep, Tnominalize are suitable defined for the application field of theme.

- **Transformation Tpass:**

Consider the terminal string:

NP1 – V – NP2

The transformation Tpass will convert this string into following strings:

NP2 – được – NP1 – V

Or,

NP2 – được – V – bởi – NP1

For example, consider the active sentence:

Tác giả A viết sách B

Terminal string of this sentence:

Tác giả A – viết – sách B

Apply Tpass, we derive:

Sách B – được – tác giả A – viết

Sách B – được – viết – bởi – tác giả A

- **Transformation Tq**

The transformation Tq will generate the class of Yes/No questions.

Consider the terminal string of affirmative sentence:

Tác giả A – viết – sách B

Apply transformation Tq to this terminal string, we derive the following terminal string:

Tác giả A – viết – sách B

At the result, we derive the interrogative sentence:

Tác giả A viết sách B?

- **Transformation Tw**

Consider the following pair of questions:

Tác giả nào viết tài liệu X
Tác giả A viết tài liệu nào

To generate these questions, we apply the optional transformation Tw which operate on any string of the form: NP1 – V – NP2.

Transformation Tw can operate in either cases:

Case 1: $Tw_1$ converts the string of the form "NP1 – V – NP2" into the string of the form "NP1 – nào/gì – V – NP2".

Case 2: $Tw_2$ converts the string of the form "NP1 – V – NP2" into the string of the form "NP1 – V – NP2– nào/gì".

Transformation Tw can apply only to strings to which Tq has already applied.

Consider the terminal string:

Tác giả A – viết – sách B

Apply transformation Tq to this string, we derive:

Tác giả A – viết – sách B (1)

Then apply transformation Tw to the string (1) in two steps:

Choosing the noun phrase "Tác giả A", the string (1) will be transformed to (2) by $Tw_1$:

Tác giả nào – viết + sách B (2)

Choosing the noun phrase "sách B", the string (1) will be transformed to (3) by $Tw_2$:

Tác giả A + viết – sách nào (3)

- **Transformation Tobsep**

Transformation Tobsep apply to any string of the form:

X – Va – Comp – NP

Which have V → Va + Comp

Consider the sentence:

Tác giả A viết sách B năm Y

In this case, the verb part has the structure "động từ + bổ ngữ" with the verb "viết" and the complement "năm Y".

To form this sentence, we apply transformation Tobsep to the terminal string:

Tác giả A – viết năm Y – sách B

- **Transformation Tnominalize**

The set of nominalizing transformations will convert sentences into noun phrases. In this paper, we mention the class of possessive sentences.

Example sentence: Tác giả A có sách B

The terminal string of this sentence:

Tác giả A – có – sách B

The transformation Tnominalize converts this string into the following string:

Sách B – của – tác giả A

## 3. Method of determining the sequence of equivalent transformations

In our approach, the sequences of equivalent transformations are defined for each kernel sentence structure. Thus, from one affirmative kernel sentence structure, the sequences of equivalent transformations which are proper defined for this structure will convert it into interrogative sentences having the equivalent meanings.

The generation of interrogative sentences having the equivalent meanings will be illustrated in following samples:

A) Group1:

The structure of kernel sentence:

danh_ngữ_một + động_từ + danh_ngữ_hai.

Example 1: Tác giả A viết tài liệu X.

Table 2 show the statistics of sentences transformed from the kernel sentence in example 1.

Table 2: Sentences transformed from the kernel sentence in example 1

| *Transformational history* | *Sentences having the equivalent meanings* |
|---|---|
| Tq | Tác giả A viết tài liệu X?<br>Tác giả A đã viết tài liệu X?<br>Tác giả A viết tài liệu X (à / đấy à / nhỉ / ....)?<br>Tác giả A đã viết tài liệu X (à / đấy à / nhỉ / ....)?<br>(Có phải là / có đúng là) tác giả A viết tài liệu X (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tác giả A đã viết tài liệu X (không) (vậy / nhỉ / ...)?<br>Tác giả A viết tài liệu X (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tác giả A đã viết tài liệu X (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tpass<br><br>(equivalent to the sequence of transformations: Tq) | Tài liệu X được tác giả A viết?<br>Tài liệu X được viết bởi tác giả A?<br>Tài liệu X đã được tác giả A viết?<br>Tài liệu X đã được viết bởi tác giả A?<br>Tài liệu X được tác giả A viết (à / đấy à / nhỉ / ...)?<br>Tài liệu X được viết bởi tác giả A (à / đấy à / nhỉ / ...)?<br>Tài liệu X đã được tác giả A viết (à / đấy à / nhỉ / ...)?<br>Tài liệu X đã được viết bởi tác giả A (à / đấy à / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X được tác giả A viết (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X được viết bởi tác giả A (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X đã được tác giả A viết (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X đã được viết bởi tác giả A (không) (vậy / nhỉ / ...)?<br>Tài liệu X được tác giả A viết (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X được viết bởi tác giả A (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X đã được tác giả A viết (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết bởi tác giả A (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ một | Ai / tác giả nào / người nào viết tài liệu X<br>Ai / tác giả nào / người nào đã viết tài liệu X?<br>Ai / tác giả nào / người nào viết tài liệu X (vậy / nhỉ / ...)?<br>Ai / Tác giả nào / Người nào đã viết tài liệu X (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ một → Tpass<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ một) | Tài liệu X được ai / tác giả nào / người nào viết?<br>Tài liệu X được viết bởi ai / tác giả nào / người nào?<br>Tài liệu X đã được ai / tác giả nào / người nào viết?<br>Tài liệu X đã được viết bởi ai / tác giả nào / người nào?<br>Tài liệu X được ai / tác giả nào / người nào viết (vậy / nhỉ / ...)?<br>Tài liệu X được viết bởi ai / tác giả nào / người nào (vậy / nhỉ / ...)?<br>Tài liệu X đã được ai / tác giả nào / người nào viết (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết bởi ai / tác giả nào / người nào (vậy / nhỉ / ...) ? |
| Tq → Tw danh ngữ hai | Tác giả A viết tài liệu nào?<br>Tác giả A đã viết tài liệu nào?<br>Tác giả A viết tài liệu nào (vậy / nhỉ / ...)?<br>Tác giả A đã viết tài liệu nào (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai → Tpass<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ hai) | Tài liệu nào được tác giả A viết?<br>Tài liệu nào được viết bởi tác giả A?<br>Tài liệu nào đã được tác giả A viết?<br>Tài liệu nào đã được viết bởi tác giả A?<br>Tài liệu nào được tác giả A viết (vậy / nhỉ / ...)?<br>Tài liệu nào được viết bởi tác giả A (vậy / nhỉ / ...)?<br>Tài liệu nào đã được tác giả A viết (vậy / nhỉ / ...)?<br>Tài liệu nào đã được viết bởi tác giả A (vậy / nhỉ / ...)? |

B) Group 2:

The structure of kernel sentence:

danh_ngữ_một + động_từ + danh_ngữ_hai + bổ_ngữ

Example 2: Tác giả A viết tài liệu X năm Y.

Table 3 show the statistics of sentences transformed from the kernel sentence in example 2.

Table 3: Sentences transformed from the kernel sentence in example 2

| *Transformational history* | *Sentences having the equivalent meanings* |
|---|---|
| Tq | Tác giả A viết tài liệu X năm Y?<br>Tác giả A đã viết tài liệu X năm Y?<br>Tác giả A viết tài liệu X năm Y (à / đấy à / nhỉ / ....)?<br>Tác giả A đã viết tài liệu X năm Y (à / đấy à / nhỉ / ....)?<br>(Có phải là / có đúng là) tác giả A viết tài liệu X năm Y (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tác giả A đã viết tài liệu X năm Y (không) (vậy / nhỉ / ...)? |

| | |
|---|---|
| | Tác giả A viết tài liệu X năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tác giả A đã viết tài liệu X năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tpass<br><br>(equivalent to the sequence of transformations: Tq) | Tài liệu X được tác giả A viết năm Y?<br>Tài liệu X được viết bởi tác giả A năm Y?<br>Tài liệu X đã được tác giả A viết năm Y?<br>Tài liệu X đã được viết bởi tác giả A năm Y?<br>Tài liệu X được tác giả A viết năm Y (à / đấy à / nhỉ / ...)?<br>Tài liệu X được viết bởi tác giả A năm Y (à / đấy à / nhỉ / ...)?<br>Tài liệu X đã được tác giả A viết năm Y (à / đấy à / nhỉ / ...)?<br>Tài liệu X đã được viết bởi tác giả A năm Y (à / đấy à / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X được tác giả A viết năm Y (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X được viết bởi tác giả A năm Y (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X đã được tác giả A viết năm Y (không) (vậy / nhỉ / ...)?<br>(Có phải là / có đúng là) tài liệu X đã được viết bởi tác giả A năm Y (không) (vậy / nhỉ / ...)?<br>Tài liệu X được tác giả A viết năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X được viết bởi tác giả A năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X đã được tác giả A viết năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết bởi tác giả A năm Y (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tobsep → Tpass<br><br>(equivalent to the sequence of transformations: Tq) | Tài liệu X được viết năm Y bởi tác giả A?<br>Tài liệu X đã được viết năm Y bởi tác giả A?<br>Tài liệu X được viết năm Y bởi tác giả A (à / đấy à / nhỉ / ...)?<br>Tài liệu X đã được viết năm Y bởi tác giả A (à / đấy à / nhỉ / ...)?<br>(Có phải là / Có đúng là) tài liệu X được viết năm Y bởi tác giả A (không) (vậy / nhỉ / ...)?<br>(Có phải là / Có đúng là) tài liệu X đã được viết năm Y bởi tác giả A (không) (vậy / nhỉ / ...)?<br>Tài liệu X được viết năm Y bởi tác giả A (có đúng không / có phải không) (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết năm Y bởi tác giả A (có đúng không / có phải không) (vậy / nhỉ / ...)? |

| | |
|---|---|
| | nhỉ / ...)? |
| Tq → Tw danh ngữ một | Ai / tác giả nào / người nào viết tài liệu X năm Y?<br>Ai / tác giả nào / người nào đã viết tài liệu X năm Y?<br>Ai / tác giả nào / người nào viết tài liệu X năm Y (vậy / nhỉ / ...)?<br>Ai / tác giả nào / người nào đã viết tài liệu X năm Y (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ một → Tpass<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ một) | Tài liệu X được ai / tác giả nào / người nào viết năm Y?<br>Tài liệu X được viết bởi ai / tác giả nào / người nào năm Y?<br>Tài liệu X đã được ai / tác giả nào / người nào viết năm Y?<br>Tài liệu X đã được viết bởi ai / tác giả nào / người nào năm Y?<br>Tài liệu X được ai / tác giả nào / người nào viết năm Y (vậy / nhỉ / ...)?<br>Tài liệu X được viết bởi ai / tác giả nào / người nào năm Y (vậy / nhỉ / ...)?<br>Tài liệu X đã được ai / tác giả nào / người nào viết năm Y (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết bởi ai / tác giả nào / người nào năm Y (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ một → Tobsep → Tpass<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ một) | Tài liệu X được viết năm Y bởi ai / tác giả nào / người nào?<br>Tài liệu X đã được viết năm Y bởi ai / tác giả nào / người nào?<br>Tài liệu X được viết năm Y bởi ai / tác giả nào / người nào (vậy / nhỉ / ...)?<br>Tài liệu X đã được viết năm Y bởi ai / tác giả nào / người nào (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai | Tác giả A viết tài liệu nào năm Y?<br>Tác giả A đã viết tài liệu nào năm Y?<br>Tác giả A viết tài liệu nào năm Y (vậy / nhỉ / ...)?<br>Tác giả A đã viết tài liệu nào năm Y (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai → Tpass<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ hai) | Tài liệu nào được tác giả A viết năm Y?<br>Tài liệu nào được viết bởi tác giả A năm Y?<br>Tài liệu nào đã được tác giả A viết năm Y?<br>Tài liệu nào đã được viết bởi tác giả A năm Y?<br>Tài liệu nào được tác giả A viết năm Y (vậy / nhỉ / ...)?<br>Tài liệu nào được viết bởi tác giả A năm Y (vậy / nhỉ / ...)?<br>Tài liệu nào đã được tác giả A viết năm Y (vậy / nhỉ / ...)?<br>Tài liệu nào đã được viết bởi tác giả A năm Y (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai → Tobsep → Tpass<br><br>(equivalent to the | Tài liệu nào được viết năm Y bởi tác giả A?<br>Tài liệu nào đã được viết năm Y bởi tác giả A?<br>Tài liệu nào được viết năm Y bởi tác giả |

| | |
|---|---|
| sequence of transformations: Tq → Tw danh ngữ hai) | A (vậy / nhỉ / ...)?<br>Tài liệu nào đã được viết năm Y bởi tác giả A (vậy / nhỉ / ...)? |

C) Group 3:

The structure of kernel sentence:
danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng.

Example 3: Tài liệu X có tác giả là A.

Table 4 show the statistics of sentences transformed from the kernel sentence in example 3.

Table 4: Sentences transformed from the kernel sentence in example 3

| *Transformational history* | *Sentences having the equivalent meanings* |
|---|---|
| Tq | Tài liệu X có tác giả là A?<br>Tài liệu X có tác giả là A (à / đấy à / nhỉ / ....)?<br>(Có phải là / có đúng là) Tài liệu X có tác giả là A (không) (vậy / nhỉ / ...)?<br>Tài liệu X có tác giả là A (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tnominalize<br><br>(equivalent to the sequence of transformations: Tq) | Tác giả của tài liệu X là A?<br>Tác giả của tài liệu X là A (à / đấy à / nhỉ / ....)?<br>(Có phải là / có đúng là) tác giả của tài liệu X là A (không) (vậy / nhỉ / ...)?<br>Tác giả của tài liệu X là A (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai | Tài liệu X có tác giả là ai?<br>Tài liệu X có tác giả là ai (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai → Tnominalize<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ hai | Tác giả của tài liệu X là ai?<br>Tác giả của tài liệu X là ai (vậy / nhỉ / ...)? |

D) Group 4:

The structure of kernel sentence:
danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai.

Example 4: Tác giả A có tài liệu X.

Table 5 show the statistics of sentences transformed from the kernel sentence in example 4.

Table 5: Sentences transformed from the kernel sentence in example 4

| *Transformational history* | *Sentences having the equivalent meanings* |
|---|---|
| Tq | Tác giả A có tài liệu X?<br>Tác giả A có tài liệu X (à / đấy à / nhỉ / ....)?<br>(Có phải là / Có đúng là) tác giả A có tài liệu X (không) (vậy / nhỉ / ...)?<br>Tác giả A có tài liệu X (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tnominalize<br><br>(equivalent to the sequence of transformations: Tq) | Tài liệu X của tác giả A?<br>Tài liệu X của tác giả A (à / đấy à / nhỉ / ....)?<br>(Có phải là / có đúng là) tài liệu X của tác giả A (không) (vậy / nhỉ / ...)?<br>Tài liệu X của tác giả A (có đúng không / có phải không) (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ một | Tác giả nào có tài liệu X ?<br>Tác giả nào có tài liệu X (vậy / nhỉ / ...) ? |
| Tq → Tw danh ngữ một → Tnominalize<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ một) | Tài liệu X của tác giả nào?<br>Tài liệu X của tác giả nào (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai | Tác giả A có tài liệu nào?<br>Tác giả A có tài liệu nào (vậy / nhỉ / ...)? |
| Tq → Tw danh ngữ hai → Tnominalize<br><br>(equivalent to the sequence of transformations: Tq → Tw danh ngữ hai) | Tài liệu nào của tác giả A?<br>Tài liệu nào của tác giả A (vậy / nhỉ / ...)? |

## 4. Developments

We defined the transformations Tpass, Tq, Tw, Tobsep, Tnominalize for 7 kernel sentence structures (affirmative) which are presented in Table 6. The sequences of equivalent transformations for these kernel sentence structures are presented in Table 1.

Table 6: List of kernel sentence structures (affirmative)

| | *The structure of kernel sentence* | *The kernel sentences* |
|---|---|---|
| Group 1 | danh_ngữ_một + động_từ + danh_ngữ_hai | Tác giả A viết tài liệu X |
| | | Nhà xuất bản P phát hành tài liệu X |
| | danh_ngữ_một + động_từ + danh_ngữ_hai + | Tác giả A viết tài liệu X do nhà xuất bản P phát hành |

| | | |
|---|---|---|
| | định_ngữ | Nhà xuất bản P phát hành tài liệu X do tác giả A viết |
| Group 2 | danh_ngữ_một + động_từ + danh_ngữ_hai + bổ_ngữ | Tác giả A viết tài liệu X năm Y |
| | | Nhà xuất bản P phát hành tài liệu X năm Y |
| Group 3 | danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng | Tài liệu X có tác giả là A |
| | | Tài liệu X có nhà xuất bản là P |
| | danh_ngữ_một + định_ngữ + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng | Tài liệu X do nhà xuất bản P phát hành có tác giả là A |
| | | Tài liệu X do tác giả A viết có nhà xuất bản là P |
| Group 4 | danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai | Tác giả A có tài liệu X |
| | | Nhà xuất bản P có tài liệu X |
| | danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + định_ngữ | Tác giả A có tài liệu X do nhà xuất bản P phát hành |
| | | Nhà xuất bản P có tài liệu X do tác giả A viết |

We have developed the system which allows to determine the Vietnamese questions having the equivalent meanings based on the method we proposed. We tested the system with more than 898 pair of questions having the equivalent meanings and 1,000 pair of questions having the different meanings. These pairs of questions are selected from 2,260 questions transformed from 14 kernel sentences which are presented in Table 6.

## 5. Conclusions

In this research, based on transformational rules in "Syntactic Structures" theory of N. Chomsky [1], we established the method which can allows to determine the Vietnamese questions having the equivalent meanings. This method is based on the hypothesis in which we can define the sequences of equivalent transformations for determined kernel sentence structures. Thus, to determine two any Vietnamese questions having the equivalent meanings or not, the system will analyze the syntactic structure of each question to determine the transformational history and the kernel sentence of each one. If these two questions were transformed from the same kernel sentence with sequences of equivalent transformations, then they are considered equivalent meanings.

The result of our research can be developed and applied to the Vietnamese language based document retrieval system has already established in [2], [3], [4], [5], [6].

## References

[1] Noam Chomsky, Syntactic Structures, Second Edition, Mouton de Gruyter, 2002.

[2] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Semantic Model for Representing Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011), vol. 4, February 26-28, 2011, Singapore, pp. 331-335. ISBN: 978-1-4244-9252-7.

[3] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Understanding the Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011), vol. 4, February 26-28, 2011, Singapore, pp. 327-330. ISBN: 978-1-4244-9252-7.

[4] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Semantic Representation for Processing a Series of Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Computer Engineering and Applications (ICCEA 2011), vol. 2, July 15-17, 2011, Haikou, China, pp. 575-578. ISBN: 978-981-08-9196-1.

[5] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "A Semantic Approach to Answer Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 International Conference on Software Technology and Engineering (ICSTE 2011), August 12-14, 2011, Kuala Lumpur, Malaysia, pp. 314-320. ISBN-13: 978-0-7918-5979-7.

[6] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Answering a Series of Vietnamese Questions in Library Retrieval System", The 2011 The 2nd International Conference on Future Information Technology (ICFIT 2011), IPCSIT vol. 13 (2011), September 16-18, 2011, Singapore, pp. 445-449. ISSN: 2010-460X.

# Handover Priority Schemes for Multi-Class Traffic in LEO Mobile Satellite Systems

**Amr S. Matar, Gamal Abd-Elfadeel, Ibrahim I. Ibrahim and Hesham M. Z. Badr**

**Department of Communication, Electronics and Computer Engineering, University of Helwan
Cairo, Egypt**

**Abstract**

In this paper, an analytical framework is developed to evaluate the performance of complete sharing (CS) with two different handover priority schemes for multi-class traffic in Low Earth Orbit-Mobile Satellite Systems (LEO-MSS).
In the first priority scheme, the handover requests are given the higher priority using queuing scheme with also taking into consideration the priority between classes of traffic. Where, in the second priority scheme a combination of guard channel and queuing of handover requests scheme is developed.

***Keywords:*** *Complete Sharing, Queuing, Multi-class traffic, LEO.*

## 1. Introduction

The great advance in technology in the last few years made it possible to use satellite networks as the backbone for wireless personal communication services (PCS) to meet third generation (3G) requirements of providing multimedia services, such as video-on-demand, multimedia games [1].
In recent years, LEOs and GEOs (Geostationary Earth Orbit satellites) have been used in commercial MSSs to directly provide voice and data services to handheld mobile terminals (MTs). Examples of such systems include the Iridium, the Globalstar (LEO–MSSs) [3], [4], and the Thuraya and the Asia Cellular Satellite (GEO–MSSs) [5], [6]. While fewer numbers of geostationary satellites are needed to cover the globe than low earth orbiting (LEO) satellites, a geostationary satellite network has larger propagation delays and requires more power for transmission than that experienced in LEO satellite networks. As a result, LEO satellites are better suited for providing real- time interactive and multimedia services than geostationary satellites [7].
To achieve efficient frequency reuse, the satellite footprint (which is a circular area on the earth surface) is divided into smaller cells or spotbeams [2]. Two different schemes are proposed regarding cellular coverage geometry for LEO satellites: (a) Satellite Fixed Cell (SFC) systems, and (b) Earth Fixed Cell (EFC) systems [8]. As most of the research works on handover schemes in space networks are carried out on Satellite Fixed Cell (SFC) systems, this paper deals with the second system.
The central issue in defining resource management strategies for LEO-MSS system is to select the suitable policy for managing handover requests. From the user standpoint, the interruption of a conversation is more undesirable than blocking of a newly arriving call. Previous researches have considered various resource management strategies for LEO-MSS. One approach is to reserve resources before handover occurrences in order to minimize forced termination probability [9, 10]. Another approach for managing handovers is to queue handover (QH) requests [11, 12]. In this approach, the queuing of handover requests is set to a maximum time interval in case there is no channel available in the destination cell. The call will be forced termination if no channel is made available within the defined time limit. This technique avoids protracted reservation of resources and favors low blocking probability but it introduces relatively high forced termination probability if the acceptable queuing delay is low. In the previous approaches, single class traffic was considered. For multi-class traffic, the performance analysis of a complete sharing (CS) with fixed channel reservation is considered in [14].
In this paper, we present an analytical framework for evaluating the performance of LEO-MSS multiclass traffic using complete sharing (CS) with two different handover priority schemes. The queuing of handover requests scheme is developed first. Second, a combination of guard channel and handover request queuing approach is examined. The results are compared with the handover priority scheme developed in [14].
This paper is organized as follows: Section 2 deals with the basic assumptions. Section 3 presents a suitable mobility model. Queuing time statistics are presenting in Section 4. An analytical study for the CS with the two priority schemes is presented in Section 5. Finally, Section 6 deals with the analytical results for the performance analysis.

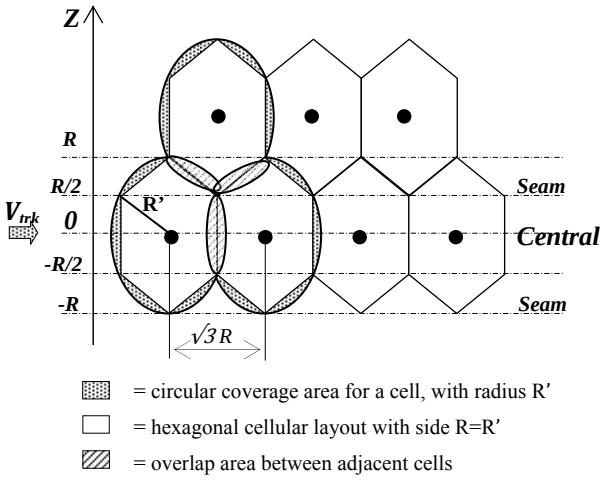



Fig. 1. The geometry assumed for the overlap areas (hexagonal cell side = circular coverage radius).

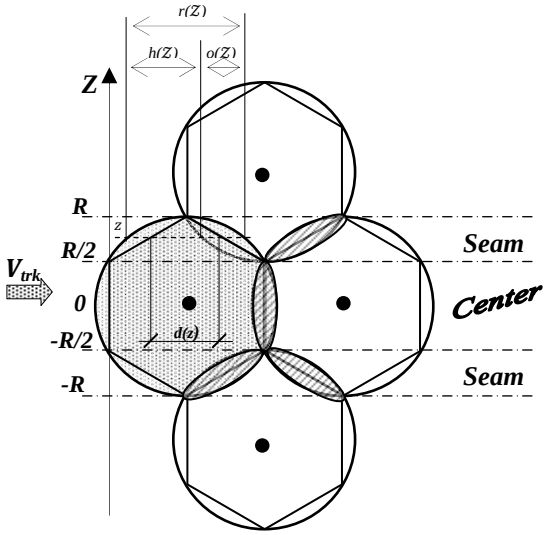



Fig. 2. The shape of the curvilinear cell and the distance crossed in the cell in the overlap area for a given height z.

## 2. Basic Assumption

Due to beam-forming, spot-beams are disposed on the earth according to a hexagonal regular layout (side R) with circular coverage of radius *R'* .The possible values for the ratio $R`/R$ range from 1 to 1.5 [13]. Clearly, the greater this ratio is, the larger the overlap area (between adjacent cells) as shown in Fig. 1. Let us assume minimum possible extension for the overlap area such that $R`=R$. The centers of adjacent cells are separated by a distance equal to $\sqrt{3}$R.
This paper is based on IRIDIUM system, but the results obtained are generally valid for all LEO-MSS's based on moving cells approach. In the IRIDIUM case, the radius, *R*, equal to 212.5 km with 66 satellites orbiting over six near polar circular orbits at about 780 km of altitude.
Assume a system serving multi-class traffic where the following quality of service (QoS) parameters [14] is used to evaluate the performance of channel resource management strategies:
1) $P_{bk}$, blocking probability of class-*k* new call attempts, representing the average fraction of new class-*k* calls that are not admitted into LEO-MSS because of unavailability of channels;
2) $P_{fk}$, handover failure probability of class-*k* calls, representing the average fraction of handover attempts of the class-*k* calls that are unsuccessful;
3) $P_{dk}$, call dropping probability of class-*k* calls, representing the average fraction of new class-*k* calls that are not blocked but eventually forced into termination due to the handover failure;
4) $P_{usk}$, unsuccessful call probability of class-*k* traffic, representing the fraction of new class-*k* calls that are not completed because of either being blocked initially or being dropped due to the failure of subsequent handover requests.
Based on ITU-T recommendations for land mobile services [15], the values $P_{bk}$ and $P_{dk}$ should not exceed $5.10^{-4}$,$10^{-2}$ respectively. Also if they may seem too severe, we consider that these requirements will be valid for future high-quality MSS's.

## 3. Mobility Model

First, we point out that this model takes into account that an MS with a class-k traffic may cross the cellular layout not only along the central region of cells (see Fig. 1), but also through the seam of the cellular network [17]. In such a case, we expect that the number of inter-beam handovers during call lifetime is significantly increased which is more realistic evaluation of the impact of user mobility on the performance of channel allocation techniques for LEO-MSS's.
In the following, we define source cell: the cell where the MS call starts and transit cell: any subsequent cell reached by the MS with the call in progress. Referring to a given cell x, with the subscript $i$ = 1 refers to the statistical parameters related to calls started in cell x, whereas subscript $i$ = 2 refers to the parameters related to handed-over calls to cell x.
The high value of the satellite ground-track speed, $V_{trk}$ (about 26600 km/h in the LEO case) with respect to the other motions such as earth's rotation around its axis and the user's motion relative to the earth, the relative satellite-user motion will be approximated by the vector $V_{trk}$. Moreover, mobile stations (MS's) cross the cellular network irradiated by a satellite according to a parallel

straight lines. The proposed model for LEO mobility is based on the following assumptions [17]:

1) MS's cross the cellular network with a relative velocity (i.e., vector $V_{trk}$), disposed as shown in Fig. 1 with respect to the cellular layout.
2) When a handover occurs, the destination cell will be the neighboring cell in the direction of the relative satellite-MS motion.
3) MS's cross the cellular network with an offset uniformly distributed all over the network.
4) From the call arrival in a cell, a random offset $z \in [-R, R]$ is associated to this call, where z is the offset of the related MS according to the reference shown in Fig. 2. the related MS travels a distance in this cell which is :
   - Uniformly distributed between zero and d(z), if the cell is the source cell of the call;
   - Deterministically equal to d(z), if the cell is a transit cell of the call.

where

$$d(z) = \begin{cases} \sqrt{3}R & if\ |z| \le \frac{R}{2} \\ 2\sqrt{3}(R - |z|) & if\ \frac{R}{2} \le |z| \le R \end{cases} \quad (1)$$

In order to characterize the user's (relative) mobility for class-$k$ traffic in LEO-MSS's we introduce the dimensionless parameter $\alpha_k$ as:

$$\alpha_k = \frac{\sqrt{3}R}{V_{trk}T_{dk}} \quad (2)$$

where
$T_{dk}$ is the average duration time of class-k calls.

Based on [17], the handover probabilities of class-$k$ traffic $P_{H1K}$, $P_{H2K}$ : are expressed as

$$P_{H1k} = \frac{2}{3}\left\{P_{h1k} + \frac{1-P_{h1k}}{\alpha_k}\right\} \quad (3)$$

$$P_{H2k} = \frac{P_{h1k}+P_{h2k}}{2} \quad (4)$$

where

$$P_{h1k} = \frac{1-e^{-\alpha_k}}{\alpha_k} \quad , \quad P_{h2k} = e^{-\alpha_k} \quad (5)$$

The channel holding time for calls in cell x [16]:

$$t_{Hik} = min[t_{dk}, t_{mci}], \quad i = 1,2. \quad (6)$$

with expected value [17]:

$$E_k[t_{Hik}] = T_{dk}(1 - P_{Hik}), \quad i = 1,2. \quad (7)$$

## 4. Analysis of Queuing Time

Let us assume that an active MS moves from cell x toward an adjacent cell y. There is an area where this MS can receive a signal with an acceptable power level from both cells; this is the so-called overlap area. The MS crosses the overlap area in a time $t_{w\,max}$.
The position of the MS at the call arrival instant is defined, as offset $z$, is assigned to this MS in the source cell. According to the basic assumptions (see Section II) and the mobility model (see Section III), the randomness of $t_{w\,max}$ only depends on the offset z assigned to the call in its source cell; in particular, $t_{w\,max}$ is derived as the time spent by the associated MS to cross the overlap area at a given offset $z$ (See Fig. 2) with a speed $V_{trk}$.

$$t_{w\,max} = \frac{O(z)}{V_{trk}} \quad (8)$$

where
$O(z)$ is the distance covered by the MS in the overlap area, which due to both the regular cellular layout and the mobility assumptions, it remains the same for any handover request. Let r(z) denote the distance coverd by the MS in the circular cell of radius R at offset z (see Fig.2).

$$r(Z) = 2\sqrt{R^2 - Z^2} \quad (9)$$

The circular cell is divided into two regions: 1) the overlap area with adjacent cells in the direction of the satellite-user's relative motion and 2) the remaining part of the cell that is called curvilinear cell. The curvilinear cell (whose area is equal to $3\sqrt{3}R^2/2$) is not hexagonal, but it is represented by the shaded area in Fig. 2. h(z) has denoted by the distance crossed by the MS in the curvilinear cell at a height z, and O(z) has denoted by the relevant distance covered in the overlap area.

$$h(z) = r(z) - O(z) \quad (10)$$

$$O(z) = \begin{cases} 2\sqrt{R^2 - Z^2} - \sqrt{3}R, & \text{if } |z| \le \frac{R}{2} \\ \sqrt{R^2 - Z^2} - \frac{\sqrt{3}}{2}R + \sqrt{R^2 - (|z| - \frac{2}{3}R)^2}, & \text{if } \frac{R}{2} \le |z| \le R \end{cases} \quad (11)$$

According to [17], the average value of the maximum queuing time $E[t_{w\,max}]$ results is:

$$E[t_{w\,max}] = \frac{E[O(z)]}{V_{trk}} = \alpha_k T_{dk} \beta \quad (12)$$

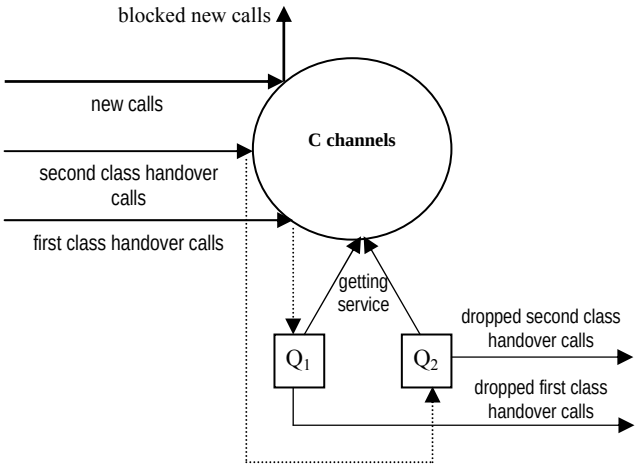



Fig. 3 Handover request queuing scheme model.

where β is given by:

$$\beta = \frac{4}{9}\left(\frac{\sqrt{3}}{3}\pi - \frac{3}{2}\right) \approx 0.1394 \qquad (13)$$

From (12), the average value of the maximum queuing time $E[t_{w\,max}]$ depends on system parameters such as the speed, cell size and does not depend on the traffic class.

## 5. Complete Sharing Performance Analysis

In this section, analytical approaches for evaluating the CS performance with two different handover priority schemes for multi-class traffic are presented. In performing our analysis, we have assumed the following:

- C channels are assigned per cell.
- New call arrivals and handover attempts of class-k traffic are two independent Poisson processes, with mean rates $\lambda_{nk}$ and $\lambda_{hk}$ respectively. And with $\lambda_{hk}$ related to $\lambda_{nk}$ by [17]:

$$\frac{\lambda_{hk}}{\lambda_{nk}} = \frac{2}{3}(1-P_{bk})\left\{\frac{P_{h1k}}{1-(1-P_{fk})P_{h2k}} + \frac{1-P_{h1k}+(1-P_{fk})(P_{h1k}-P_{h2k})}{\alpha_k - \alpha_k(1-P_{fk})^2 P_{h2k}}\right\} \qquad (14)$$

- Whether class-k handover requests are queued or not, the channel holding time in a cell (for both new call arrivals and handovers) is approximated by a random variable with an exponential distribution and mean $1/\mu_k$ given by [17]:

$$\frac{1}{\mu_k} = \frac{\lambda_{nk}(1-P_{bk})}{\lambda_{nk}(1-P_{bk})+\lambda_{hk}(1-P_{fk})}E_k[t_{H1k}] + \frac{\lambda_{hk}(1-P_{fk})}{\lambda_{nk}(1-P_{bk})+\lambda_{hk}(1-P_{fk})}E_k[t_{H2k}] \qquad (15)$$

- The maximum waiting time is approximated by a random variable exponentially distributed, with expected value equal to $1/\mu_w = E[t_{w\,max}]$ , where $E[t_{w\,max}]$ is given by (12).

### 5.1 Complete Sharing (CS) with Handover Request Queuing Scheme

In this subsection, an analytical approach to queuing of handover requests scheme is developed. The proposed queuing model is shown in fig.3. In general, when there are free channels in the cell, new calls and/or handover calls are equally likely to get service. However, when all the channels are occupied, new calls are blocked whereas handover call requests are queued in their respective queues (first class handover request is queued in its queue (Q1) of Length k and second class handover requests in (Q2) of Length L) for a maximum time $t_{w\,max}$, waiting for a free channel according to their priorities. The first class handover requests have higher priority over second class handover requests. If the queues are full, handover calls are dropped.

Let $\Lambda(j)$ denotes the number of free channels in the generic cell $j$. According to this queuing scheme, the inter-beam handover requests are as follows:

1) If $\Lambda(j) \neq \emptyset$, the new and handover calls get service immediately in cell j.
2) If $\Lambda(j) = \emptyset$, the new calls are blocked and the handover requests are queued waiting for an available channel in cell $j$. In the meantime, the handover call is served by its originating cell. A handover request leaves the queue for one of the following reasons:
   a) The handover procedure is successful: The handover request is served, before the call is ended and its maximum queuing time has expired.
   b) The handover procedure has been useless: The call ends before the corresponding handover request is served and its maximum queuing time has expired.
   c) The handover procedure fails and the call is dropped.

According to the queuing scheme described, the queuing scheme can be modeled as an M/M/C/K queue. The evolution of queue can be described by the Markov chain in Fig. 4.

From the two-dimension (2-D) Markov chain shown in Fig.4, Let us define $S_{n,i,j}$ as the state of the cell where $n$ is the number of busy channels (first and second classes, new and handoff call) and $i$ and $j$ signifies the number of handover call requests of first and second classes in queue Q1 and the queue Q2 respectively. The transition between states can be explained as follows:

• A transition from state $S_{n,0,0}$ to $S_{n+1,0,0}$ for $0 \leq n < C$ occurs when a new call or handover call (either class one or class two) arrives, thus it occurs with rate $\lambda = \lambda_n + \lambda_h$

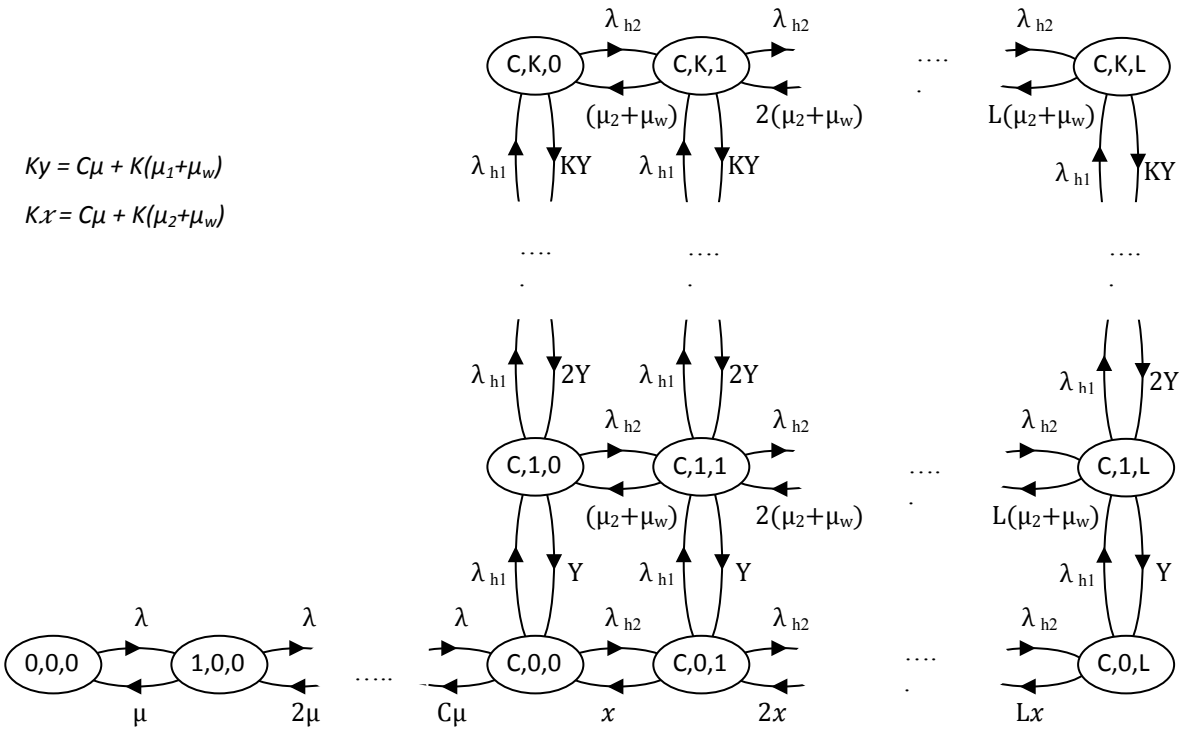



Fig.4. Markov chain representation of the CS with handover requests queuing priority scheme.

(where $\lambda_n$ is the total new call arrival rate $\{\lambda_{n1} + \lambda_{n2}\}$, and $\lambda_h$ is the total handover call arrival $\{\lambda_{h1} + \lambda_{h2}\}$).

• A transition from state $S_{n,0,0}$ to state $S_{n-1,0,0}$ for $0 < n \leq C$ occurs if a call in progress finishes its service and releases the channel, thus occurs with rate $n\mu$ (where $\mu$ is the total call departure rate which equal to $\{\mu_1 + \mu_2\}$).

• When all channels are busy, a transition to the next states occurs if there is a first or second-class handover call arrival *and* the first or second-class queue is not full.

Hence, a transition from state $S_{C,i,j}$ to state $S_{C,i+1,j}$ occurs with rate $\lambda_{h1}$, while a transition from state $S_{C,i,j}$ to state $S_{C,i,j+1}$ occurs with rate $\lambda_{h2}$.

• A transition from state $S_{C,i,j}$ to state $S_{C,i-1,j}$occurs if a channel is released *and* the first-class handover call gets service *or* the first-class handover call finishes its call while in the queue, *or* the waiting time in the queue for a handover call is over before a channel is released, thus occurs with rate $C\mu + i(\mu_1 + \mu_w)$.

• A transition from state $S_{C,i,j}$ to state $S_{C,i,j-1}$ occurs if the waiting time for a second-class handover call is over before a channel is released *or* the second-priority handover call finishes its call while in the queue, *or* a channel is released and a second-class handover call gets served provided there is no handover call waiting in first-class handover queue, thus it occurs with rate or with rate $C\mu + i(\mu_2 + \mu_w)$.

Based on the above descriptions and Fig. 4, the Balance equation describing this model:

$$\lambda P_{n,i,j} = (n+1)\mu P_{n+1,i,j}, \quad i = 0, j = 0, 0 \leq n < C \tag{16}$$

$$(\lambda_h + C\mu)P_{n,i,j} = \lambda P_{c-1,i,j} + YP_{c,i+1,j} + \mathrm{XP_{c,i,j+1}}, \quad i = 0, j = 0, n = C \tag{17}$$

$$(\lambda_h + iY)P_{c,i,j} = \lambda_{h1}P_{c,i-1,j} + (i+1)YP_{c,i+1,j} + (\mathrm{j}+1)\left(\mu_2 + \mu_w\right)\mathrm{P_{c,i,j+1}}, \quad 0 < i < K, j = 0, n = C \tag{18}$$

$$(\lambda_{h2} + iY)P_{c,i,j} = \lambda_{h1}P_{c,i-1,j} + (\mathrm{j}+1)\left(\mu_2 + \mu_w\right)\mathrm{P_{c,i,j+1}}, \quad i = K, j = 0, n = C \tag{19}$$

$$(\lambda_h + j\mathrm{X})P_{c,i,j} = \lambda_{h2}P_{c,i,j-1} + (j+1)\mathrm{X}P_{c,i,j+1} + (\mathrm{i}+1)\mathrm{YP_{c,i+1,j}}, \quad i = 0, 0 < j < L, n = C \tag{20}$$

$$(\lambda_{h1} + j\mathrm{X})P_{c,i,j} = \lambda_{h2}P_{c,i,j-1} + (\mathrm{i}+1)\mathrm{YP_{c,i+1,j}}, \quad i = 0, j = L, n = C \tag{21}$$

$$\left(\lambda_{h2} + iY + j\left(\mu_2 + \mu_w\right)\right)P_{c,i,j} = \lambda_{h1}P_{c,i-1,j} + \lambda_{h2}P_{c,i,j-1} + (\mathrm{j}+1)\left(\mu_2 + \mu_w\right)\mathrm{P_{c,i,j+1}}, \quad i = K,\ 0 < j < L, n = C \tag{22}$$

$$\left(\lambda_{h1} + iY + j\left(\mu_2 + \mu_w\right)\right)P_{c,i,j} = \lambda_{h1}P_{c,i-1,j} + \lambda_{h2}P_{c,i,j-1} + (i+1)YP_{c,i+1,j}, \quad 0 < i < K, j = L, n = C \tag{23}$$

$$\left(\lambda_h + iY + j\left(\mu_2 + \mu_w\right)\right) P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1} + (i+1) Y P_{c,i+1,j} + (j+1)\left(\mu_2 + \mu_w\right) P_{c,i,j+1}, \quad 0 < i < K, 0 < j < L, n = C \qquad (24)$$

$$\left(iY + j\left(\mu_2 + \mu_w\right)\right) P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1} \quad i = K, j = L, n = C \qquad (25)$$

The steady-state probabilities $P_{n,i,j}$ that the cell is in state $S_{n,i,j}$ can be found by solving the previous balance equations and the normalization condition $\sum_{n=0}^{C} \sum_{i=0}^{K} \sum_{j=0}^{L} P_{n,i,j} = 1$.

New call blocking occurs if a new call arrival (either class one or class two) finds C channel occupied. Therefore, the steady state blocking probability for the new calls can be expressed as

$$P_{B1} = P_{B2} = \sum_{i=0}^{K} \sum_{j=0}^{L} P_{c,i,j} \qquad (26)$$

where

$P_{B1}$ and $P_{B2}$ are the new call blocking probabilities for class one and two respectively.

Handover failure occurs if a handover call arrival finds all channels are occupied and its respective request queue is full *or* the handover call request is queued in its respective queue; however, it is dropped before getting service because its waiting time in the queue is expired before the handover call gets served or finished its service.

The steady-state handover failure probability of class-one traffic is given as

$$P_{F1} = \sum_{j=0}^{L} P_{C,K,j} + \sum_{i=0}^{K-1} \sum_{j=0}^{L} P_{f1;i,j} P_{C,i,j} \qquad (27)$$

where the first term describes the event that the first-class handover request queue is full. While the While second term describes the event that the first-class handover call request is queued, but it is dropped before getting service because its waiting time is expired before a channel is released. The term $P_{f1;i,j}$ gives the probability of handover failure for a first-class handover call request in the queue given the handover call request joined the queue as the (i+1) call. This is found as [18]:

$$P_{f1;i,j} = \frac{(i+1)\mu_w}{C\mu + i(\mu_1 + \mu_w)} \qquad (28)$$

Similar the steady-state handover failure probability of class-two traffic is given as:

$$P_{F2} = \sum_{i=0}^{K} P_{C,i,L} + \sum_{j=0}^{L-1} \sum_{i=0}^{K} P_{f2;i,j} P_{C,i,j} \qquad (29)$$

where the first term describes the event that the second-class handover request queue is full. While the second term describes the event that the second-class handover call request is queued, but it is dropped before getting service because its waiting time is expired before a channel is released. The term $P_{f2;i,j}$ gives the probability of handover failure for a second-class handover call in the queue given the handover call joined the queue as the ( $j$+1) call. This obtained as:

$$P_{f2;i,j} = \frac{(j+1)\mu_w}{C\mu + j(\mu_2 + \mu_w)} \qquad (30)$$

The probability of an admitted call being forced into termination during the $i^{th}$ handover can be expressed as

$$P_{dki} = P_{Fk}\left[P_{h1k}(1 - P_{Fk})^{i-1} P_{h2k}^{i-1}\right] \qquad (31)$$

By summing over all possible values of $i$, $P_{dk}$ can be obtained as follows

$$P_{dk} = \sum_{i=1}^{\infty} P_{dki} = \sum_{i=1}^{\infty} P_{Fk}\left[P_{h1k}(1 - P_{Fk})^{i-1} P_{h2k}^{i-1}\right] = \frac{P_{Fk} P_{h1k}}{1 - P_{h2k}(1 - P_{Fk})} \qquad (32)$$

$P_{usk}$ is also used as an important parameter for evaluating overall system performance and can be derived as

$$P_{usk} = P_{Bk} + P_{dk}(1 - P_{Bk}) \qquad (33)$$

## 5.2 Complete Sharing (CS) with Guard Channel and Queuing of Handover Requests Scheme

This subsection presents an analytical model for the combination of guard channel and handover request queuing scheme. In this model, when there are free channels in the cell, new calls and/or handover calls are equally likely to get service. However, When the number of occupied channels are equal to threshold (M=C-$C_h$), new calls are blocked whereas handover calls are gets service. When all the channels are occupied, handover call requests are queued in their respective request queues (first class handoff call is queued in its queue (Q1) of Length k and second class requests in (Q2) of Length L) for a maximum time $t_{w\,max}$, waiting for a free channel according to the same scenario discussed in the previous scheme.

This scenario can be represented by the second-Dimension (2-D) Markov chain shown in Fig. 5. Let us define $S_{n,i,j}$ as the state of the cell where n is the number of busy channels (first and second classes, new and handoff call) and i and j signifies the number of handover call requests of first and second classes in queue Q1 and the queue Q2 respectively. The transition between states can be explained as follows:

• A transition from state $S_{n,0,0}$ to $S_{n+1,0,0}$ for $0 \le n < M$ occurs when a new call or handover call (either class one or class two) arrives, thus it occurs with rate $\lambda = \lambda_n + \lambda_h$ (where $\lambda_n$ is the total new call arrival rate $\{\lambda_{n1} + \lambda_{n2}\}$, and $\lambda_h$ is the total handover call arrival $\{\lambda_{h1} + \lambda_{h2}\}$).

• A transition from state $S_{n,0,0}$ to $S_{n+1,0,0}$ for $M \leq n < C$ occurs when a handover call (either class one or class two) arrives, thus it occurs with rate $\lambda_h$.

• A transition from state $S_{n,0,0}$ to state $S_{n-1,0,0}$ for $0 < n \leq C$ occurs if a call in progress finishes its service and releases the channel, thus occurs with rate $n\mu$ (where $\mu$ is the total call departure rate which equal to $\{\mu_1 + \mu_2\}$).

• When all channels are busy, a transition to the next states occurs if there is a first or second-class handover call arrival *and* the first or second-class queue is not full.

Hence, a transition from state $S_{C,i,j}$ to state $S_{C,i+1,j}$ occurs with rate $\lambda_{h1}$, while a transition from state $S_{C,i,j}$ to state $S_{C,i,j+1}$ occurs with rate $\lambda_{h2}$.

• A transition from state $S_{C,i,j}$ to state $S_{C,i-1,j}$ occurs if a channel is released *and* the first-class handover call gets service *or* the first-class handover call finishes its call while in the queue, *or* the waiting time in the queue for a handover call is over before a channel is released, thus occurs with rate $C\mu + i(\mu_1 + \mu_w)$.

• A transition from state $S_{C,i,j}$ to state $S_{C,i,j-1}$ occurs if the waiting time for a second-class handover call is over before a channel is released *or* the second-priority handover call finishes its call while in the queue, *or* a channel is released and a second-class handover call gets served provided there is no handover call waiting in first-class handover queue, thus it occurs with rate or with rate $C\mu + i(\mu_2 + \mu_w)$.

Based on the above descriptions and Fig. 5, the Balance equation describing this model:

$$\lambda P_{n,i,j} = (n+1)\mu P_{n+1,i,j}, \quad i = 0, j = 0, 0 \leq n < C - C_h \qquad (34)$$

$$\lambda_h P_{n,i,j} = (n+1)\mu P_{n+1,i,j}, \quad i = 0, j = 0, C - C_h \leq n < C \qquad (35)$$

$$(\lambda_h + C\mu)P_{n,i,j} = \lambda_h P_{c-1,i,j} + YP_{c,i+1,j} + \mathrm{XP_{c,i,j+1}}, \quad i = 0, j = 0, n = C \qquad (36)$$

$$(\lambda_h + iY)P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + (i+1)YP_{c,i+1,j} + (\mathrm{j}+1)(\mu_2 + \mu_w)\mathrm{P_{c,i,j+1}}, \quad 0 < i < K, j = 0, n = C \qquad (37)$$

$$(\lambda_{h2} + iY)P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + (\mathrm{j}+1)(\mu_2 + \mu_w)\mathrm{P_{c,i,j+1}}, \quad i = K, j = 0, n = C \qquad (38)$$

$$(\lambda_h + j\mathrm{X})P_{c,i,j} = \lambda_{h2} P_{c,i,j-1} + (j+1)\mathrm{X}P_{c,i,j+1} + (\mathrm{i}+1)\mathrm{YP_{c,i+1,j}}, \quad i = 0, 0 < j < L, n = C \qquad (39)$$

$$(\lambda_{h1} + j\mathrm{X})P_{c,i,j} = \lambda_{h2} P_{c,i,j-1} + (\mathrm{i}+1)\mathrm{YP_{c,i+1,j}}, \quad i = 0, j = L, n = C \qquad (40)$$

$$\left(\lambda_{h2} + iY + j(\mu_2 + \mu_w)\right)P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1} + (\mathrm{j}+1)(\mu_2 + \mu_w)\mathrm{P_{c,i,j+1}}, \quad i = K, 0 < j < L, n = C \qquad (41)$$

$Ky = C\mu + K(\mu_1 + \mu_w)$

$Kx = C\mu + K(\mu_2 + \mu_w)$

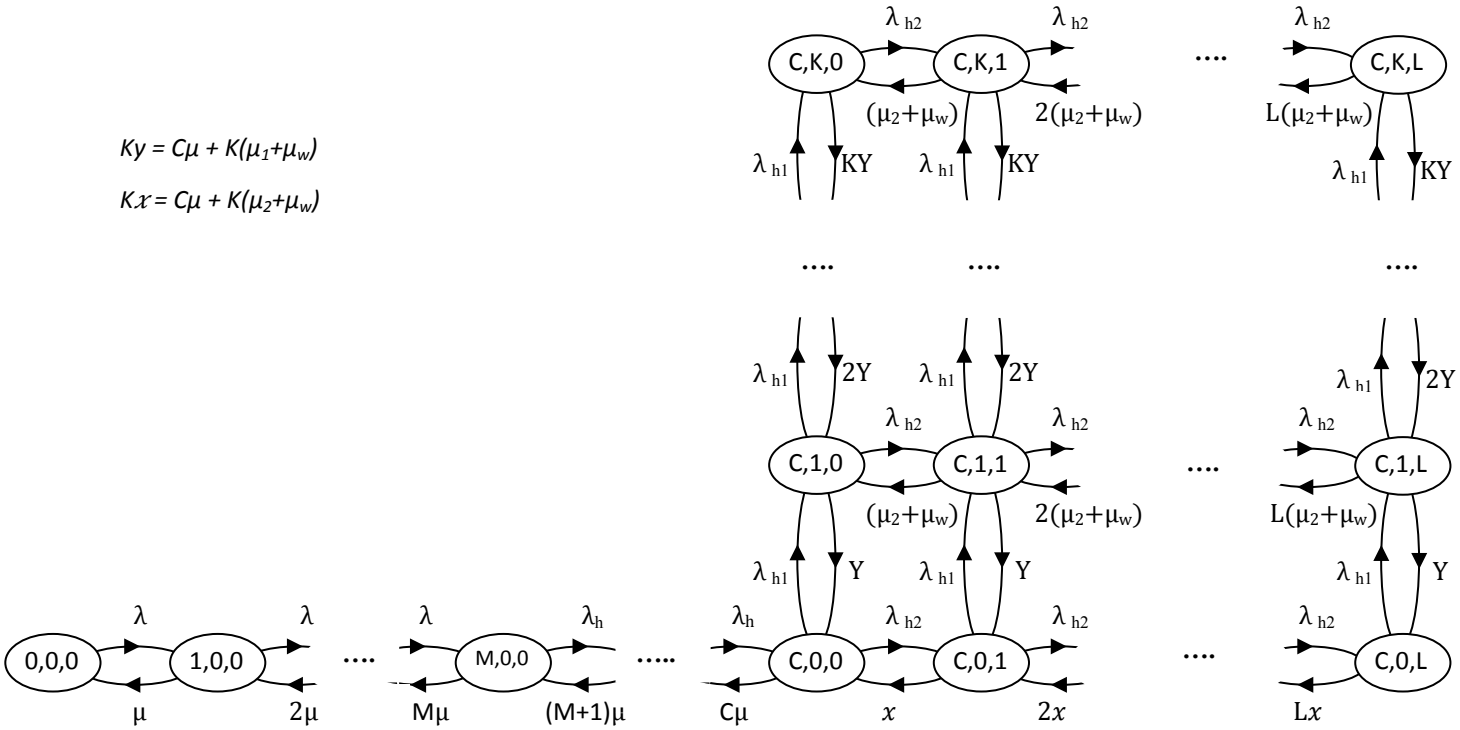
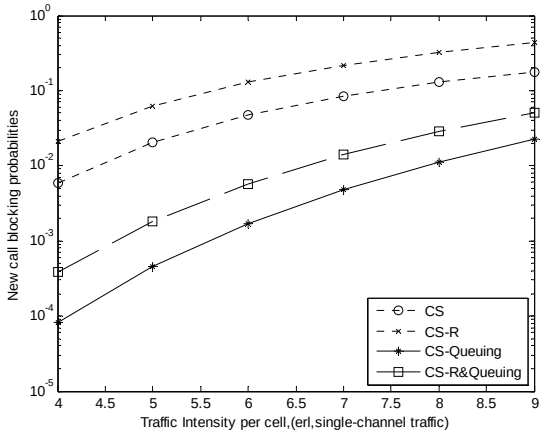
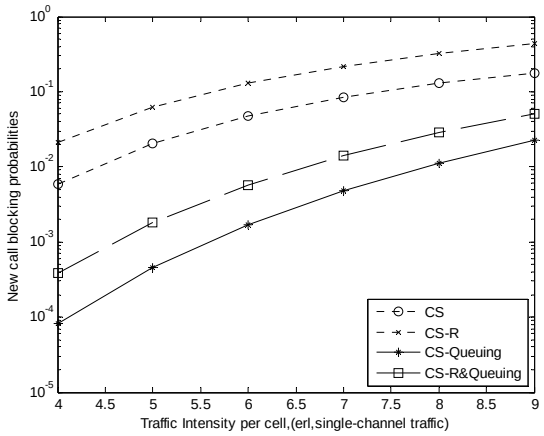



Fig.5 . Markov chain representation of the CS with combination of guard channel and handover request queuing priority scheme.

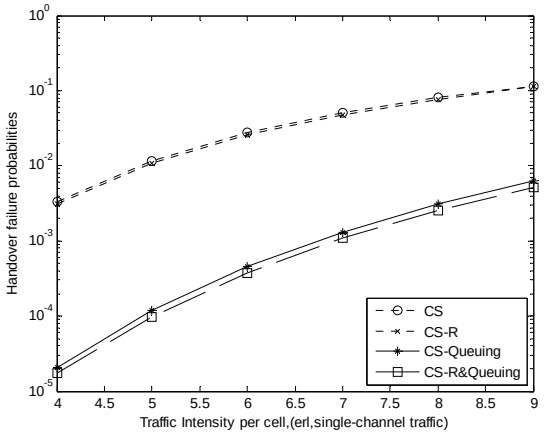



(a)

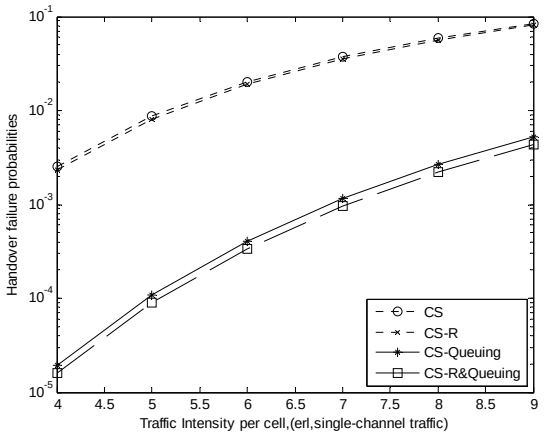



(b)

Fig. 6. Analytical results for new call blocking probabilities as function of traffic intensity of CS with different priority schemes. (a) first class. (b) second class.




(a)




(b)

Fig. 7. Analytical results for Handover failure probabilities as function of traffic intensity of CS with different priority schemes. (a) first class. (b) second class.

$$\left(\lambda_{h1} + iY + j(\mu_2 + \mu_w)\right) P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1} + (i+1) Y P_{c,i+1,j}$$
$$0 < i < K, j = L, n = C \qquad (42)$$

$$\left(\lambda_h + iY + j(\mu_2 + \mu_w)\right) P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1} + (i+1) Y P_{c,i+1,j} + (j+1)(\mu_2 + \mu_w) P_{c,i,j+1},$$
$$0 < i < K, 0 < j < L, n = C \qquad (43)$$

$$\left(iY + j(\mu_2 + \mu_w)\right) P_{c,i,j} = \lambda_{h1} P_{c,i-1,j} + \lambda_{h2} P_{c,i,j-1}$$
$$i = K, j = L, n = C \qquad (44)$$

The steady-state probabilities $P_{n,i,j}$ that the cell is in state $S_{n,i,j}$ can be found by solving the previous balance equations and the normalization condition $\sum_{n=0}^{C} \sum_{i=0}^{K} \sum_{j=0}^{L} P_{n,i,j} = 1$.

New call blocking occurs if a new call arrival ( either class one or class two ) finds ($C$-$C_h$) channel occupied. Therefore, the steady state blocking probability for the new calls can be expressed as:

$$P_{B1} = P_{B2} = \sum_{n=C-C_h}^{C-1} P_{n,0,0} + \sum_{i=0}^{K} \sum_{j=0}^{L} P_{c,i,j} \qquad (45)$$

where

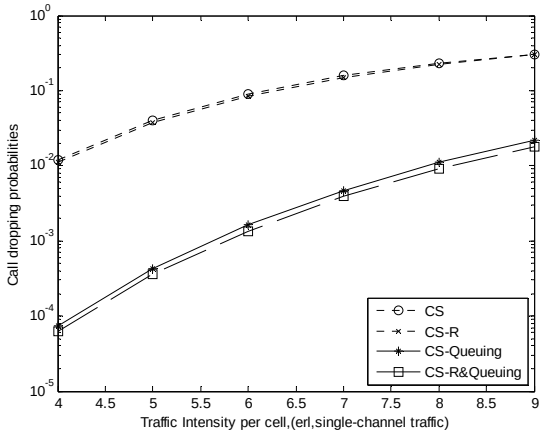
$P_{B1}$ and $P_{B2}$ are the new call blocking probabilities for traffic of class one and class two respectively.

Handover failure occurs if a handover call arrival finds all channels are occupied and its respective request queue is full *or* the handover call request is queued in its respective queue; however, it is dropped before getting service because its waiting time in the queue is expired before the handover call gets served or finished its service.

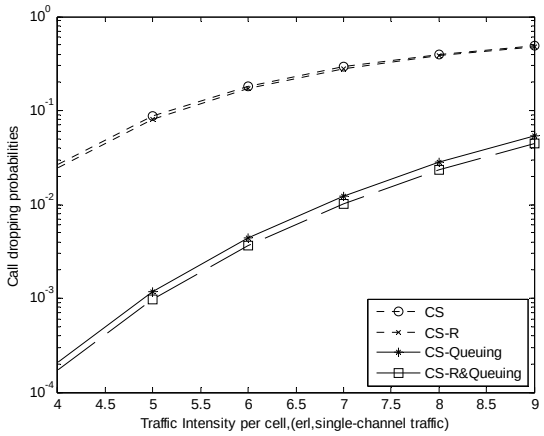
The steady-state handover failure probability of class-one traffic is given as

$$P_{F1} = \sum_{j=0}^{L} P_{C,K,j} + \sum_{i=0}^{K-1} \sum_{j=0}^{L} P_{f1;i,j} P_{C,i,j} \qquad (46)$$

where the first term describes the event that the first-class handover request queue is full, while the second term describes the event that the first-class handover call request is queued, but it is dropped before getting service because its waiting time is expired before a channel is

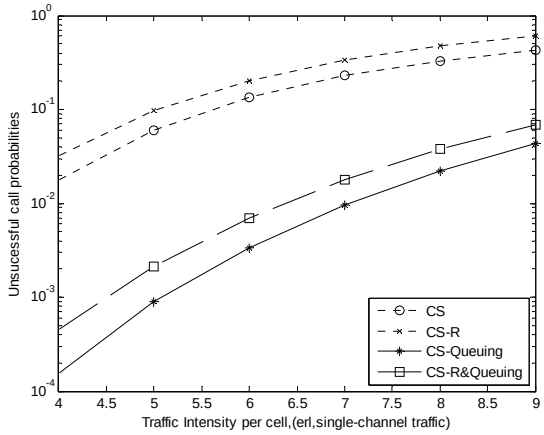



(a)

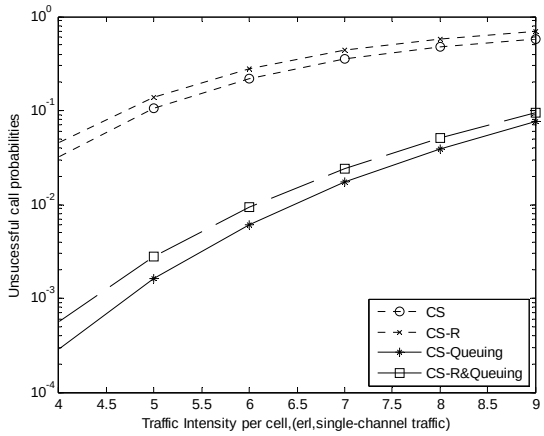



(b)

Fig. 8. Analytical results for call dropping probabilities as function of traffic intensity of CS with different priority schemes. (a) first class. (b) second class.




(a)




(b)

Fig. 9. Analytical results for Unsuccessful call probabilities as function of traffic intensity of CS with different priority schemes. (a) first class. (b) second class.

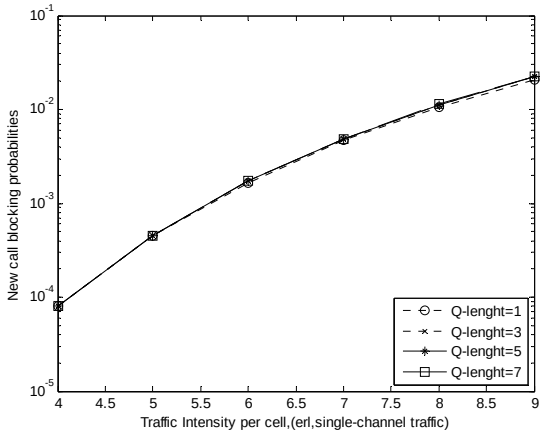
released. The term $P_{f1;i,j}$ gives the probability of handover failure for a first-class handover call request in the queue given the handover call request joined the queue as the ($i$+1) call. This is found as [18]:

$$P_{f1;i,j} = \frac{(i+1)\mu_w}{C\mu+i(\mu_1+\mu_w)} \quad (47)$$

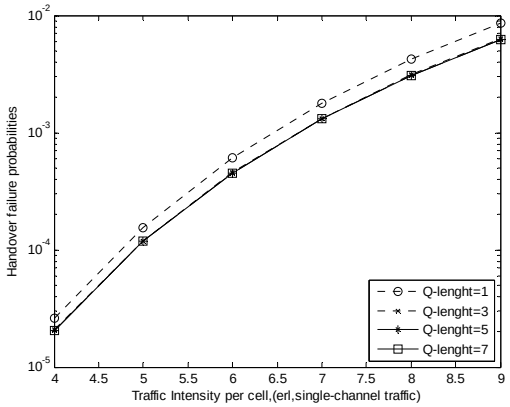
Similar the steady-state handover failure probability of class-two traffic can be computed as (29).
Using (32) and (33), $P_{dk}$ and $P_{usk}$ can then be computed, respectively.

## 6. Analytical Result

The main goal of this section is to analyze the analytical results of the CS with handover queuing priority (named as CS-Queuing) scheme and the CS with the combination of guard channel and handover queuing priority (named as CS-R&Queuing) scheme which have been presented in section **V**. The following parameter values of two different class of traffic have been chosen in the validations: $C$=10, $T_{d1}$=180, $T_{d2}$=540, $\rho 2$=0.02$\rho 1$, the first and second class of handover request queues are L=5 and K=5 respectively and channel reservation $M$=2.
Figs. 6-9 show analytical results of CS policy under different priority schemes in terms of $P_{bk}$, $P_{fk}$, $P_{dk}$ and $P_{usk}$ respectively. In these graphs, the behavior of CS with no priority (named as CS) and CS with fixed channel reservation (guard channel) priority (named as CS-R) scheme examined in [14] have been also considered.
As can be seen from Fig. 6-9, the handover queuing (CS-Queuing) and combination of handover queuing with guard channel (CS-R&Queuing) schemes provide significantly better results in terms of all quality of service parameters considered ($P_{bk}$, $P_{fk}$, $P_{dk}$ and $P_{usk}$ respectively) when compared with CS with no priority (CS) or with fixed channel reservation (CS-R) schemes[14].
In Fig. 6 the analytical results for new call blocking probability show that the handover queuing (CS-Queuing)

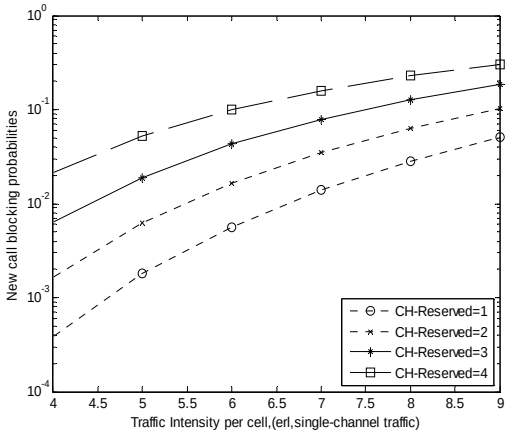



(a)

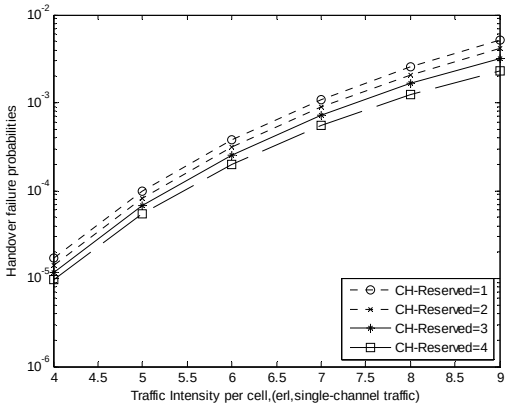



(b)

Fig. 10. The effect of Q1 length on the First class traffic :
a)New call blocking Probability. b)Handover failure Probability.




(a)




(b)

Fig. 11. The effect of channel reservation value on the First class traffic :
a)New call blocking Probability. b)Handover failure Probability.

scheme achieves better response than the handover queuing with guard channel combination (CS-R&Queuing) scheme. However, the CS-R&Queuing scheme is a little better in response of handover failure probability as shown in Fig. 7, and also in the response of call dropping probability (see Fig. 8). But as seen in Fig. 8, the unsuccessful call probability ($P_{usk}$ ) of CS-Queuing scheme is the best response over other priority schemes.

For CS-Queuing priority scheme and as we can see in Fig.10, the increasing of class-one handover request queuing ($Q_1$) length has a approximately the same effect on the response of new call blocking probability and handover failure probability of class-one traffic.

In the CS-R&Queuing priority scheme, the new call blocking probability increases significantly as the number of channel reservation increase. However, it results with a decrease in the handover failure probability as can be seen in Fig. 11

## 7. Conclusions

In this paper, we have developed an analytical work to evaluate the performance of CS resource management strategies for multi-class traffic in LEO-MSS. Two different handover priority schemes have been introduced: the handover queuing scheme and the combination of handover queuing with guard channel scheme.

Analytical results have shown that the CS with queuing of handover requests scheme effectively reduces new call blocking probability and the unsuccessful call probability with a little increase in handover failure probability than did the combination handover queuing with guard channel scheme. Therefore, CS with the handover request queuing scheme should be preferred to the combination scheme one.

## References

[1] G. Albertazzi, GE. Corazza, M. Neri, and A. Vanelli-Coralli, "Performance of turbo coding for satellite UMTS multimedia broadcast multicast services", In Proceedings of IEEE ICCT2003, Vol. 2, Apr. 2003, pp.1078-1081.

[2] M. Karaliopoulos, K. Narenthiran, B. Evans, P. Henrio, M.Mazzella, W. De Win, M.Dieudonné, P. Philippopoulos, D.I. Axiotis, I. Andrikopoulos, I.Mertzanis, G.E. Corazza, A.Vanelli-Coralli, N.Dimitriou, and A.Polydoros, "Satellite radio interface and radio resource management strategy for the delivery of multicast/broadcast services via an integrated satellite-terrestrial system", IEEE Communications Magazine, Vol. 42, No. 9, Sept. 2004, pp.108-117.

[3] 2011. [Online]. Available: http://www.iridium.com

[4] 2011. [Online]. Available: http://www.globalstar.com

[5] 2011. [Online]. Available: http://www.thuraya.com

[6] 2011 . [Online].Available: http://www.acesinternational.com

[7] S. Kalyanasundaram, E.K.P. Chong, and N.B. Shrof, "An Efficient Scheme to Reduce Handoff Dropping in LEO Satellite Systems". Kluwer Academic Publishers, Wireless Networks, Vol. 7, Issue 1, 2001, pp.75-85.

[8] J. Restrepo and G. Maral, "Coverage concepts for satellite constellations providing communications services to fixed and mobile users", Space Communications, Vol. 13, No. 2, 1995, pp.145-157.

[9] E. Papapetrou and F.-N. Pavlidou, "QoS Handover Management in LEO/MEO Satellite Systems", Wireless Personal Communications, Vol.24, No.2, 2003, pp.189-204.

[10] E. Papapetrou, S. Karapantazis, G. Dimitriadis, and F.-N. Pavlidou, "Satellite Handover Techniques for LEO Networks", International Journal on Satellite Communications and Networking, Vol. 22, No. 2, March/April 2004, pp.231–245.

[11] Z. Wang and P. Mathiopoulos, "On the Performance Analysis of Dynamic Channel Allocation with FIFO Handover Queuing in LEO-MSS", IEEE Trans. Communications, Vol. 53, No. 9, 2005, pp.1443-1446.

[12] E. Del Re, R. Fantacci, and G. Giambene, "Different queuing policies for handover requests in low earth orbit mobile satellite systems", *IEEE Transactions on Vehicular Technology,* Vol.48, No.2, Mar. 1999, pp.448-458.

[13] T. P. Chu and S. S. Rappaport, "Overlapping coverage and channel rearrangement in microcellular communication systems", IEE Proc. Commun., Vol. 142, No. 5, Oct. 1995, pp. 323–332.

[14] Z.Wang, P.T.Mathiopoulos, and R.Schober, "Channeling Partitioning Policies for Multi-Class Traffic in LEO-MSS", IEEE Trans. Aerospace and electronic systems, Vol. 45, NO. 4, Oct. 2009, pp.1320-1332.

[15] ITU-T Recommendation E.771, "Network grade of service parameters and target values for circuit-switched land mobile services".

[16] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized handoff procedures", IEEE Trans. Veh. Technol., Vol. 35, Aug. 1986, pp.77-92.

[17] E. Del Re, R. Fantacci, and G. Giambene, "Handover queuing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems", IEEE Trans. Commun., Vol. 47, No. 1, 1999, pp.89-102.

[18] A. E. Xhafa, and O. K. Tonguz, "Dynamic Priority Queuing of Handover Calls in Wireless Networks: An Analytical Framework", IEEE J. Select. Areas Commun., Vol. 22, No. 5, 2004, pp. 904-916.

# HCTE: Hierarchical Clustering based routing algorithm with applying the Two cluster heads in each cluster for Energy balancing in WSN

Nasrin Azizi[1], Jaber Karimpour[2], Farid Seifi[3]
[1]Technical and Engineering Dept., Zanjan Branch, Islamic Azad University, Zanjan, Iran
[2]Faculty of Electrical & Computer Engineering, University of Tabriz, Tabriz, Iran
[3]Technical and Engineering Dept., Tabriz Branch, Islamic Azad University, Tabriz, Iran

***Abstract*—In wireless sensor networks, the energy constraint is one of the most important restrictions. With considering this issue, the energy balancing is essential for prolonging the network lifetime. Hence this problem has been considered as a main challenge in the research of scientific communities. In the recent papers many clustering based routing algorithms have been proposed to prolong the network lifetime in wireless sensor networks. But many of them not consider the energy balancing among nodes. In this work we propose the new clustering based routing protocol namely HCTE that cluster head selection mechanism in it is done in two separate stages. So there will be two cluster head in a cluster. The routing algorithm used in proposed protocol is multi hop. Simulation Results show that the HCTE prolongs the network lifetime about 35% compared to the LEACH.**

***Keywords-wireless sensor network; multi hop routing; clustering; energy balancing***

## I. INTRODUCTION

Wireless sensor networks are containing of the thousands or more sensors that are widely distributed in the environment. Distribution of the sensors in the environment can be done manually or randomly. These networks have many applications such as environmental monitoring, healthcare, military operations, target tracking and etc. Due to the power constraint, energy balancing and maximizing network lifetime have been important challenge in wireless sensor networks. For this reason, use of data aggregation which limits redundant transmission between sensors is essential. One of the techniques that use the data aggregation to reduce energy consumption in WSNs is called clustering based routing algorithm [1-4]. In [5-10] there are protocols that use the clustering technique in the network.

LEACH is one of the most famous clustering based routing protocols in WSN [6]. Cluster head selection among sensor nodes is done randomly and also data transmitting between cluster heads and base station is done directly in the LEACH. Although this specification of LEACH avoids energy hole problem but causes the energy of cluster heads that are far from the base station be discharge faster than others.

HEED [9] is another well-known clustering based routing algorithms in WSN. Cluster head selection algorithm is based on a relationship between remaining energy and reference energy in HEED.

In this paper we propose the HCTE protocol that is a combination of new cluster head selection and routing algorithms. The cluster head selection algorithm in HCTE is done in two separate stages. In addition, data transmitting between cluster heads and base station is multi hop in proposed protocol.

The rest of the paper organized as follows: in section 2, we describe the related works. Section 3 explores the proposed algorithm with detailed. Section 4 explain the simulation parameters and result analysis. Final section is containing of conclusion.

## II. RELATED WORKS

As mentioned above in previous papers have suggested many protocols for clustering. In this section we explain the some celebrated clustering protocols.

MCBT [11] proposes a distributed algorithm to create a stable backbone by selecting the nodes with higher energy or degree as the cluster heads. LNCA [12] introduces a novel clustering algorithm which uses the similarity of sensed data as an important factor in cluster formation. In [13], a cluster-based routing protocol for wireless sensor networks with non-uniform node distribution is proposed, which includes an energy-aware clustering algorithm EADC and a cluster-based routing algorithm. EADC uses competition range to construct clusters of even sizes. In [14], the authors propose a mobility-based clustering (MBC) protocol for wireless sensor networks with mobile nodes. In this clustering protocol, a sensor node elects itself as a cluster-head based on its residual energy and mobility. A non-cluster-head node aims at its link stability with a cluster head during clustering

according to the estimated connection time. HEED [9] periodically selects cluster-heads. In this protocol, cluster-head selection is primarily based on the residual energy of each node. HEED also consider intra-cluster "communication cost" as a secondary clustering parameter.

Clustering algorithm in HEED is done in three phases that is described in more. In first phase, each of nodes identifies its neighbors in cluster range and then computes the required energy for communication with them and also calculates the primary cluster head selection probability by (1).

$$CH_{prob} = C_{prob} * (E_r / E_m). \quad (1)$$

Where $C_{prob}$ *is the* initial percentage of cluster-heads among all n nodes, $E_r$ is the estimated current residual energy in the node and $E_m$ is a reference maximum energy.

In the second phase each of nodes sends an advertisement message to the other nodes of its cluster and also receives the same message from other candidates. This advertisement message is containing of the obtained values from the first phase. Each of these candidates cancels candidacy if its cost is more than the other candidates.

In the last phase, non-cluster head nodes attempt to select a cluster head and join to it.

LEACH is containing of four phase that are the advertisement, cluster formation, scheduler creation and data transmission. In first phase, nodes compete with each other for election as cluster head, so that all nodes produce a random number between 0 and 1, then produced number be compared with threshold value which is achieved from (2). If produced number is smaller than T(n), then the node is selected as a cluster head.

$$T(n) = \begin{cases} \dfrac{p}{1 - p(r \bmod (1/p))} & n \in G \\ 0 & others \end{cases} \quad (2)$$

In this formula, p is percent of cluster heads per all nodes, r is current round and G is set of nodes that are not selected as cluster head in 1/p of last rounds. As can be inference, cluster heads election in LEACH is random operation.

In the second phase, nodes join to their near cluster head and forms the clusters. In the next phase, cluster heads create the scheduler such as TDMA. In the last phase, all nodes transmit the data to their cluster heads based on the created scheduler and cluster heads also aggregate the data before directly sending to the base station. This action will cause unbalanced energy consumption among the cluster heads and therefore lifetime of some nodes is much less than the lifetime of entire network.

## III. PROPOSED PROTOCOL

As mentioned above, HCTE is the clustering based routing algorithm that has two cluster heads namely initial and second cluster heads in each cluster and is based on multi-hop transmitting mechanism in the data routing from the cluster heads to sink. Each of these cluster heads has separate tasks in the cluster.

First, we explain the definition and assumptions used in HCTE. Then we will describe the HCTE protocol.

### A. Definition

A cluster head is a high level one if its distance to sink is less than the distance of sender cluster head to sink

### B. Assumptions

- All nodes are randomly distributed.
- Nodes are static or pseudo static.
- The initial energy is the same for all nodes.
- Nodes are aware of the location (by GPS or other positioning algorithms).
- Nodes are able to control their energy consumption.
- Cluster heads are aware from their remaining energy and also from the remained energy of their high level cluster heads.

### C. Energy Consumption Model

In HCTE, energy model is obtained from [6] that use both of the open space (energy dissipation $d^2$) and multi path (energy dissipation $d^4$) channels by taking amount the distance between the transmitter and receiver. So energy consumption for transmitting a packet of l bits in distance $d$ is given by (3).

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs} d^2 & , d \le d_0 \\ lE_{elec} + l\varepsilon_{mp} d^4 & , d > d_0 \end{cases} \quad (3)$$

In here $d_0$ is the distance threshold value which is obtained by (4), $E_{elec}$ is required energy for activating the electronic circuits. $\varepsilon_{fs}$ and $\varepsilon_{mp}$ are required energy for amplification of transmitted signals to transmit a one bit in open space and multi path models, respectively.

$$d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}. \quad (4)$$

Energy consumption to receive a packet of $l$ bits is calculated by (5).

$$E_{Rx}(l) = lE_{elec} \quad . \quad (5)$$

### D. HCTE Protocol

HCTE like other clustering based routing algorithms has several phases to configure the network in cluster form and to send the data to sink. For this reason, HCTE has five phases that are initial cluster head announcement, cluster formation, second cluster head announcement, schedule creation and data transmission.

*1) Initiall Cluster Head Announcement*

The initial cluster head announcement phase which is the initial cluster head selection phase is almost like a cluster head selection algorithm in HEED but the difference is that in the beginning all nodes calculates the probability of initial cluster head selection by (6) and then follows from the operations in the HEED.

$$C\_V_{ICH} = \alpha(\frac{E_r}{E_i}) + \beta(\frac{N_{non}}{N_{on}}) + \lambda(\frac{\sum E_{i\_non} - \sum E_{r\_non}}{\sum E_{i\_non}}) \quad (6)$$

Here, $E_r$ is remaining energy of sensor node and $E_m$ is the initial energy of sensor. $N_{non}$ of a node is the number of neighboring ordinary nodes which is in its transmission radio range and $N_{on}$ is the number of all ordinary nodes in the network. $E_{r\text{-}non}$ is remaining energy of neighboring ordinary node and $E_{i\text{-}non}$ is its initial energy. Parameters $\alpha$, $\beta$ and $\lambda$ determine the weight of each ratio so that sum of them is 1.

In fact, the node is selected as a initial cluster head that in addition to having the high level of residual energy, the number of nodes in its neighborhood was more and the average residual energy of its neighboring nodes is low.

Initial cluster heads are used for cluster formation, data gathering from cluster members and sending the data after gathering to second cluster heads in the clusters.

*2) Cluster Formation*

In cluster formation phase, each of the nodes tries to find the best cluster head and then joins to it. For this reason, the all nodes calculate the confidence value of initial cluster heads that are on their radio transmission range by (7). Then they join to the initial cluster head that its confidence value is greater than other.

$$M\_V_{ICH} = \frac{E_N + E_{ICH}}{(D_{N\_ICH})^2} \quad (7)$$

In here $E_{ICH}$ is remaining energy of initial cluster head and $E_N$ is remaining energy of node. $D_{N\_ICH}$ is the distance between desired node and the initial cluster head.

With considering this function, a node will join to the initial cluster head that in addition to having high levels of residual energy, it is also close to desired node.

If the node not able to found the cluster head on its radio transmission range, same as LEACH it join to the its near initial cluster head.

*3) Second Cluster Head Announcement*

In second cluster head announcement phase which is the second cluster head selection phase, all nodes within the clusters compete with each other on their confidence value which is obtained by (8). The node within a cluster is selected as a second cluster head that its confidence value is greater than other nodes at the same cluster.

$$C\_V_{SCH} = \alpha(\frac{E_r}{E_i}) + \beta(\frac{D_m - D_{SCH\_BS}}{D_m}). \quad (8)$$

Here, $D_m$ is the maximum distance in network and $D_{SCH\_BS}$ is the distance between desired node and sink.

In short, the node is selected as a second cluster head that has high level residual energy and also it has the minimum distance to sink.

Second cluster heads are used for data forwarding to the sink. These cluster heads forward the received data from their initial cluster heads and also from their low level second cluster heads to their high level second cluster heads or directly to the sink.

Fig. 1 shows an example of a network with initial and second cluster heads in the clusters.

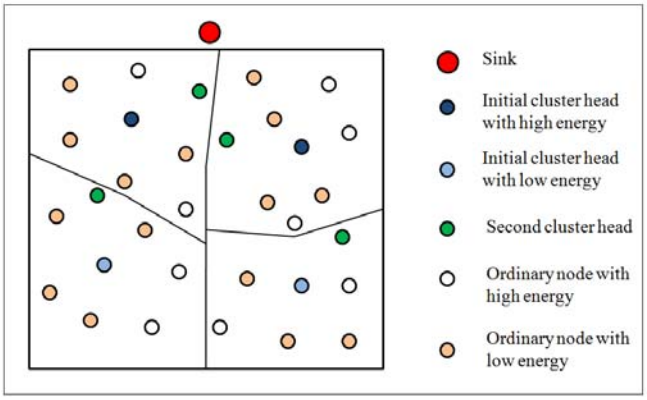



Figure 1. A network with initial and second cluster heads in the clusters

*4) Schedule Creation*

Schedule creation phase in HCTE is the same as LEACH.

*5) Data Transmission*

The other difference between LEACH and HCTE is in data transmission from cluster heads to the sink. This operation is done directly in the LEACH, but in the HCTE it is multi hop. The proposed protocol considers the distance between initial cluster heads and base station in multi hop and hence can solve the unbalanced energy consumption problem.

In data transmitting phase, in order to data sending to base station, the second cluster head must choose the best high level second cluster head with considering the several parameters such as remaining energy and distance to sink in the form of (9). This function is a cost function in the high level second cluster head selection. Hence, the second cluster head is selected among the high level second cluster heads that has the lowest cost. Fig. 2 illustrates these operations.

$$C\_F_{HCH} = \frac{(D_{CH\_HCH})^2}{E_{CH}} + \frac{(D_{HCH\_BS})^2}{E_{HCH}} \quad (8)$$

Here, $E_{CH}$ is remaining energy of source second cluster head and $E_{HCH}$ is the remaining energy of high level second cluster head. $D_{CH\_HCH}$ is the distance between source second cluster head and the high level second cluster head and $D_{HCH\_BS}$ is the distance between high level second cluster head and sink.

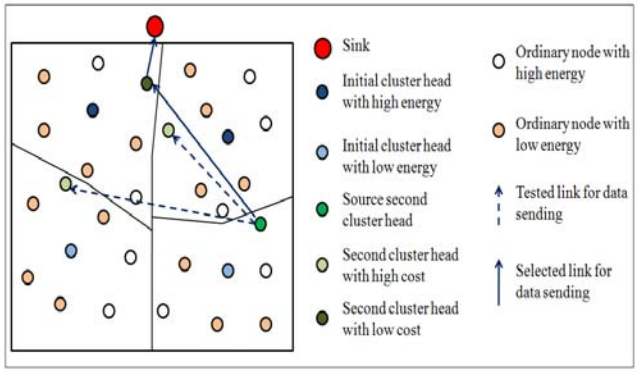
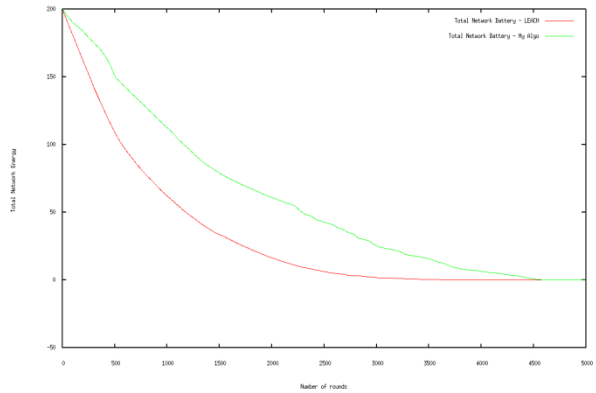



Figure 2. Illustration of data transmitting operations

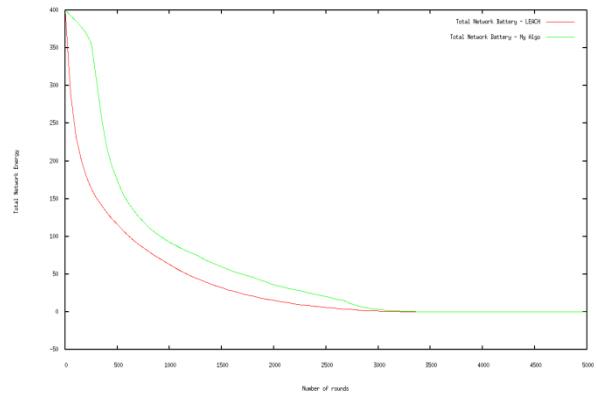
## IV. Simulation and Result Analysis

HCTE and LEACH are simulated with GCC and the simulation repeated for many times with different simulation areas and number of nodes to achieve the reliable results about proposed algorithm. Simulation parameters are presented in Table I and obtained results are shown below.

Fig. 3 (a) and 3 (b) shows the residual energy of whole network per round for LEACH and HCTE in the network with the different nodes. As it can be seen, obtained values for HCTE is better than LEACH.




(a) Residual energy of whole network per round with default parameters

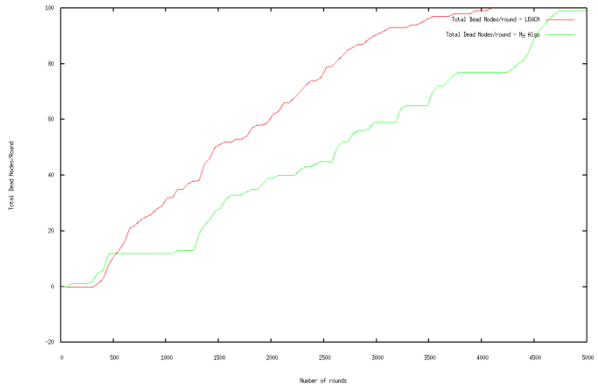



(b) Residual energy of whole network per round with 200 nodes

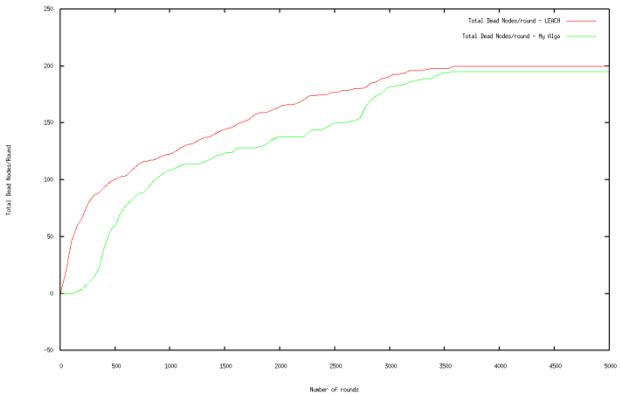
Figure 3. Residual energy of whole network per round with the different nodes in network

Fig. 4 (a) and 4 (b) shows the number of dead nodes per round in LEACH and HCTE in the network with the different nodes.

As it can be seen, proposed protocol has a performance better than LEACH protocol so that in HCTE, the times of first node dies (FND) and half nodes alive (HNA) and last node dies (LND) are optimized about 8%, 72% and 24%, respectively compared to the LEACH.




(a) Number of dead nodes per round with default parameters




(b) Number of dead nodes per round with 200 nodes

Figure 4. Number of dead nodes per round with the different nodes in network

## V. Conclusion

This work proposes a new clustering based routing protocol namely HCTE for wireless sensor networks that has two cluster heads namely initial and second cluster heads in each cluster and is based on multi-hop transmitting mechanism in the data routing from the cluster heads to sink. Each of the cluster heads has separate tasks in the cluster. The routing algorithm used in proposed protocol is multi hop so that can balance the energy consumption among nodes. Simulation results show that the HCTE prolongs the network lifetime about 35% in comparison to the LEACH.

## REFERENCES

[1] C. R. Lin, and M. Gerla, "Adaptive clustering for mobile wireless network," IEEE JOURNAL on Selected Areas in Communications, Vol. 15, No. 7, Sep. 1997, pp. 1265-1275.

[2] T. C. Hou, and T. J. Tsai, "Distributed Clustering for Multimedia Support in Mobile Multihop Ad Hoc Network", IEICE Transactions on Communications, Vol. E84-B, No. 4, April 2001, pp. 760-770.

[3] J. H. Ryu, S. H. Song, and D. H. Cho, "Energy-Conserving Clustering Scheme for Multicasting in Two-tier Mobile Ad- Hoc Networks", IEEE Electronic Letters, Vol. 37, No. 20, August 2001, pp. 1253-1255.

[4] Q. Jiang, and D. Manivannan, "Routing Protocols for Sensor Networks," IEEE Consumer Communications and Networking Conference, 2004, pp. 93-98.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An Application- Specific Protocol Architecture for Wireless Microsensor Networks," IEEE Transactions on Wireless Communications, vol. 1, no. 4, 2002.

[6] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proceedings of the Hawaii International Conference on System Sciences, 2000.

[7] O. Younis, and S. Fahmy, "Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach," Proceedings of IEEE INFOCOM, vol. 1, 2004, pp. 629-640.

[8] S. Soro, and W. B. Heinzelman, "Prolonging the Lifetime of Wireless Sensor Networks via Unequal Clustering," Proceedings of the International Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks, 2005.

[9] O. Younis, and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," IEEE Transactions on Mobile Computing, vol. 3, no. 4, 2004, pp. 366-379.

[10] S. Rasouli Heikalabad, A. Habibizad Navin, M. K. Mirnia, S. Ebadi, and M. Golesorkhtabar, "EBDHR: Energy Balancing and Dynamic Hierarchical Routing algorithm for wireless sensor networks," IEICE Electron. Express, 2010, vol. 7, no. 15, pp. 1112-1118.

[11] I. Shin, M. Kim, M. W. Mutka, H. Choo, and T. J. Lee, "MCBT: Multi-Hop Cluster Based Stable Backbone Trees for Data Collection and Dissemination in WSNs," MDPI Sensors, vol. 9, 2009, pp. 6028-6045.

[12] D. Xia, and N. Vlajic, "Near-optimal Node Clustering in Wireless Sensor Network for Environment Monitoring," 21st International Conference on Advanced Networking and Applications, 2007, pp. 632-641.

[13] Jiguo Yua, Yingying Qi, Guanghui Wang, Xin Gua, "A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution", International Journal of Electronics and Communications (AEU), Vol. 66, Iss. 1, January 2012, pp. 54-61.

[14] Deng S., Li J., Shen L., "Mobility-based clustering protocol for wireless sensor networks with mobile nodes", IET Wirel. Sens. Syst., Vol. 1, Iss. 1, 2011, pp. 39-47.

# Factors Influencing ICT Adoption in Halal Transportations: A Case Study of Malaysian Halal Logistics Service Providers

**Mohd Iskandar Illyas Tan[1], Raziah Noor Razali[2] and Mohammad Ishak Desa[3]**

**[1] Halal Informatics Research Lab (HOLLISTIC), Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia**

**[2] Halal Informatics Research Lab (HOLLISTIC), Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia**

**[3] Department of Modeling and Industrial Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia**

**Abstract**

The purpose of this study is i) to investigate the factors that influence the adoption of Information and Communication Technology (ICT) in Halal transportations and logistics and ii) to develop an ICT adoption framework for Halal logistic service providers (LSPs). The Halal LSPs selected for the study currently used ICT service platforms, such as accounting and management system for Halal logistic business. The study categorizes the factors influencing the adoption decision and process by LSPs into four groups: technological related factors, Halal assurance related factors, organizational and environmental related factors. The major contribution in this study is the discovery that technological related factors (ICT compatibility with Halal requirement) and Halal assurance related factors are the most affecting factors among the Halal LSPs applying ICT for Halal performances control in transportation's operation. Among the environmental related factors, ICT requirement for monitoring Halal included in Halal Logistic Standard on Transportation (MS2400:2010) are the most influencing factors in the adoption of ICT with the support of the government. In addition, the government related factors are very important in the reducing the main barriers and the creation of the atmosphere of ICT adoption in Halal LSP sector.

***Keywords:*** *Information Communication Technology (ICT), Halal logistic standard, Halal transportation, ICT Adoption*

## 1. Introduction

Logistics plays a key role in protecting the Halal status of any given product through proper transportation, storage and handling within the supply chain, until it reaches its final destination [1]. The main success of the Halal industry relies heavily on logistics service management capabilities in ensuring the integrity of Halal products. The logistic service management involves the collection, consolidation, storage handling, value added, track and trace and controls of the movement and storage of Halal products. For these purposes, LSPs play a crucial role in realizing this goal for the Halal industry. Segregation goods from Halal and non-Halal goods for cross-contamination avoidance are the main element of protecting the Halal status. Among the elements of controlling 'Halal' in logistic activities are monitoring Halal performances controls in transportations activities and the movements by any type of transportation mode must comply with the principle of Shariah [2]. There is a risk of cross contamination of Halal product with Non-Halal during transportation operations. Among the issues are sharing containers, poor visibility into what inventory is in which containers, where the container is transit, history of immediate suppliers, history of immediate maintenance and segregation allocation space between Halal and non-Halal goods in same containers increased the risk towards Halal integrity being compromised.

While concerning on this main issues, maintaining the Halal performance responsibility during transportation process also a big challenge. Jaafar, et al. [3] argued that to achieve a Halal supply chain compliance product is almost unattainable. This is because the Halal supply chain service offered by the LSP is guaranteed only when the products are in their custody. But, once the products are transferred to the custody of the other party, the chances of breakage in chain is higher when the other party is not practicing Halal supply chain. The lack of information sharing among suppliers and community is possible caused of these issues. Jaafar et al adds, this situation is more critical at the retail level especially the small retailers due to lack of control and monitoring by the responsible institution at their level.

Monitoring Halal integrity of product is very crucial; many researchers see the potential of ICT to improve the Halal services in logistic activities. According to Tierman [4], the use of ICT may increase the effectiveness and organization of the Halal supply chain. However the adoption of ICT in Halal industry is fairly new. According to Malaysia logistic Directory [5], a study by Frost & Sullivan also showed that the concept of adopting visibility technologies for security management purposes such as radio-frequency identification (RFID) and global positioning systems (GPS) is still fairly new despite the fact that the demand by logistics end-users who expressed their interest in the use of RFID and GPS as forms of logistics security management is high. The use of technology in logistics is currently focused on warehousing, bar coding and transportation management systems. At present, it is estimated that only 35% of logistics service providers are using the technologies. The low adoption of RFID system is due to the high initial set up cost and less mature of such technology across the ASEAN region [5].

Thus, the purpose of this study is to investigate the factors that influence the Halal LSPs in applying ICT for Halal controls transportations. To gain in-depth understanding, the contributing factors such as environment, organizational, government's responsibility and the Halal assurance element factors that contribute confidence and positive attitude towards Halal are being focused. Based on the analysis of the cases of Halal LSPs adopting ICT, the second purpose of this study is to propose a framework of ICT adoption for Halal LSPs. The proposed ICT adoption framework consists of four dimensions (technological related factors, Halal assurance related factors, organizational and environmental factors), which are partially based on some ICT adoption frameworks used in [6] [7] and [8] . Finally, this paper discusses the importance of government's role and the cooperation work between Halal industries and ICT industries and how to effectively provide initiatives to the LSPs to adopt and to retain ICT through bridging the findings in the study.

## 2. Literature Review: Factors Influencing of ICT Adoption of Malaysian Halal LSP's

In this literature review we highlight the factors affecting the ICT adoption process and their impact on Halal LSPs performance. As our study is focusing on Halal transportation's operation in Halal logistic, the impact of ICT on Halal performance efficiencies also being explored.

### 2.1 Technological Related Factors - *portions of Roger's (1995) model innovation diffusion*

In recent years, there have been many research efforts aimed at identifying factors and practices indicating how technological innovation may support company in practicing Halal controls and management in their logistics services. Malisa Mazlan [9] used Roger's theory to investigate the factors that may affect the ICT adoption process among JAKIM Halal certified company in Malaysia. They found that the companies have a high degree of adoption in the variables of relative advantage, compatibility, trialability, observability, image and complexity. The study highlighted a high degree of adoption in complexity variable indicates that they have a difficult and hard to adopting ICT in their business.

Despite the growing interest about ICT in Halal logistic [10] [9] and Halal transportations performances control [11] [12] [4] the field of ICT is relatively new and research on ICT towards Halal controls in transportations chain and logistic activities is limited [13]. Research on the importance of ICT for Malaysian Halal SMEs applied Rogers's theory [9] successfully discussed but little research is available regarding ICT adoption by Malaysian Halal LSPs. According to Tierman [10], LSPs play important roles to develop and ultimately controls the entire Halal logistical concept by mean conducting organization with a specialized and advanced ICT to make logistic operations transparent and controllable. Therefore, the study that discussed factors influences the adoption of ICT by LSPs is needed to determine the current status of ICT adoption level and to what extent the ICT has been applied in monitoring the Halal integrity.

### 2.2 Organizational Relate Factors

Besides focusing on particular ICT factors, the organizational and environmental factors also may impact on the process of ICT adoption in an organization. According to Rashid and Al-Qirim [6], the organizational factors collectively impact on the resources of the business in relation to adoption of ICT innovation [14]. However, the process of ICT adoption could be quite difficult for a firm because of its requirements. The willingness for adoption of ICT is, usually, associated with organizational readiness where organization must adapt with a large investment and firms may not have sufficient financial resources to support the high investment in hardware and software technology that is required [15, 16]. Therefore we would expect that organizational awareness, encouragement and readiness might influence technological innovation.

## 2.3 Environmental Related Factors

In addition to technological and organizational factors, the external environment in which a firm conducts its business will also influence the innovative capability [17]. Environmental factors provide significant forces for adoption where the issues relating to market climate and the firm's standing in the market directly influence the uptake of technology. Damanpour [18] found that environments with high uncertainties would have positive influence on the relationship between organizational structures and organizational innovation. Apart from that, competitor also could be one of the important external factors considered in ICT adoption. ICT adoption decision would influenced by the relative advantage gained by LSPs compared to their competitors. If there is no relative advantage gained by LSPs, ICT might not be adopted [19].

Governmental support is another important environmental characteristic for technological innovation. The government roles are importance in putting to order the local Halal industry to ensure Halal integrity. In 2010, the Halal Development Centre (HDC), as the Halal authority on behalf of Malaysian government, has launched many Halal programs for LSPs. The government also offering the investment tax allowance of 100% of qualifying capital expenditure incurred within a period of 5 years for Halal certified LSPs. In June 2010, the government announced the launching of Standards on Halal Logistics, MS 2400: 2010 that covers the Halal transportation aspects and the requirement of ICT for Halal controls. The standard stated, an organization shall establish and apply a traceability system that enables the identification of the inbound goods and/or cargo for the processing stages in the transportation chain services, history of immediate suppliers and the details of distribution routes for delivering Halal goods must be recorded [20]. These incentives have been seen to encourage new investments in 'Halal' logistics services for the export market and to increase the use of modern and state-of-the-art machinery, ICT and equipment in producing high quality 'Halal' services that also comply with the international standards.

## 2.4 Halal Assurance Related Factors

Shariah law is the fundamental guide in developing the Halal standard. The only different between conventional transportation and Halal transportations system is where the principle of the Shariah is being applied to the transportation chain. With the Halal certification and Halal standard established by government, LSPs and manufacturers are obliged to act responsibly to maintain the *halal* status of the Halal services they offered. To avoid the risk towards Halal being compromised, effective control measures, providing Halal assurance system need to be implemented by LSPs [1] [21]. In this way, it encourages confidence in the safety also the Halal integrity of products and thus promotes both confidence in the Halal industry and stability of Halal businesses [21]. In literature, Azah et al [22] discovered that there is no real time Halal tracking implemented by Halal LSPs. Azah adds that the issue of applying ICT towards Halal is still in early stage. According to Zailani [19], there is no method to determine whether the food product come from the country which is stated on its packaging. This finding has created opportunities to other Halal LSPs to encourage in developing an ICT solution and adoption for tracking purposes.

## 2.5 The ICT Adoption Framework

Based on review of the literature on factors that impact the adoption process, a conceptual framework was developed as shown in Figure 1. The study explored IS/IT adoption and diffusion models for LSPs and identified the essential factors that may impact the ICT adoption process by Halal LSPs. As this study focusing on Halal, we identify the Halal assurance related factors are among the factors that influence the adoption process. Four attributes from Roger's model [23] are relative advantage, compatibility, complexity, image and cost which will be adapted in this framework to test the impacts on ICT adoption process. This theory will be used to analyze technology factor. In the case of Halal study, we categorized the compatibility and relative advantage variables of ICT as the degree to which the adoption is perceived as benefits and compatible with Halal requirement jobs responsibilities and value system. While an image variable is the degree to which adopting ICT is perceived to enhance Halal LSPs image or status. Based on the literature, we categorize the factors influencing the adoption decision of Halal LSPs into four dimensions: technology, Halal assurance related issues, organizational and environmental related factors.

The framework was developed after undertaking an in- depth review of literature relating to ICT innovation diffusion, competitive advantage and also Halal factors itself. This review gave us an indication of the important questions that should be asked to LSPs to better understand the factors of ICT adoption. We developed a questionnaire consists of 8 questions that sought general information such as company name, size etc as well as 18 specific questions about ICT adoption issues. We tested this adoption framework on three case studies of Malaysian Halal LSPs.

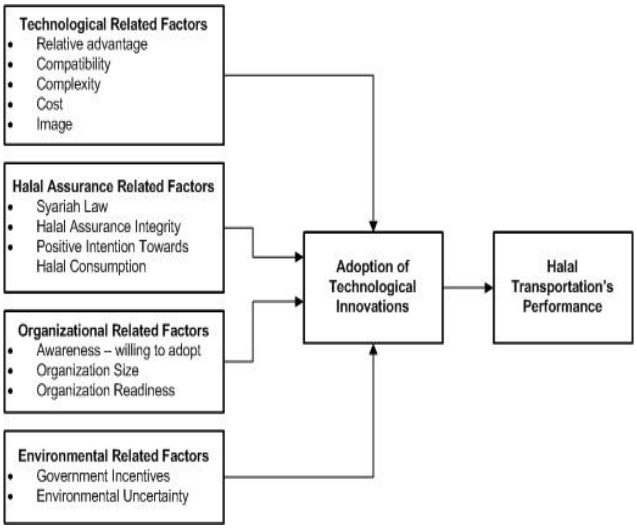



Figure 1: Research Framework – *Factors Influencing ICT Adoption in Halal Transportation by Malaysian Halal LSPs*

## 3. Research Methodology

### 3.1 Data collection

This research applies a case study methodology for data collection and analysis. The reason for choosing this methodology was to provide qualitative data that can help us better understand how ICT is initiated within the Malaysian Halal LSPs organizations and to expose factors that supported ICT diffusion. In addition, case studies help us understand the details of the cases from the participant's viewpoint by using multiple sources of data [24]. As this research focused on understanding how diffusion occurs, an exploratory case study approach was adopted. For research that was an exploratory in nature, qualitative methods were deemed more appropriate [24]. This multiple case study explores the factors that may impact the ICT adoption process and also explores the ICT application used to monitor Halal performances control in transportation's operation.

Among the research questions are;

- What are the driving factors that best support the implementation of ICT initiatives in Malaysian Halal LSP?
- What ICT characteristic suit for Halal transportation requirement that can be identified for Malaysian Halal LSP?
- What are the factors that may impact on the process of ICT adoption in Halal transportation for Halal LSP?

### 3.2 Case Study Criteria

Malaysian Halal LSPs were selected from a benchmark list in Malaysian Halal logistic industry as potential case candidates. All Halal LSPs had used ICT systems that supported communication and document management within their logistic activities and supply chain management. High-level managers (vice presidents or other high level managers) were interviewed to provided related data experience for the study. The case study interviews were conducted from April 2011 until September 2011. Case data were collected primarily through structured face-to-face interviews with managers of these Halal LSPs companies. However, when necessary, telephone interview with other executives in the firms were conducted to supplement the information gathered during the personal interviews. To enhance answer validity, participants verified the summaries of major findings of each interview after the end of each interview session. Furthermore, to ensure consistency and reliability, structured guidelines were used for all interviews.

## 4. Case Study: Finding and Data Analysis

The results were analyzed using QSR NVivo 9 software analysis. NVIVO is one of CAQDAS (computer-assisted qualitative data analysis) that can enhance the qualitative research process, quickly process queries, and expand analytical avenues [25]. Even though the individual had different software application experience from each organization, they did share over 70% of a common ICT experience in conducting Halal transportations operation.

The findings from the case studies confirmed the four main factors of ICT diffusion within Halal LSPs companies. Before discussing these factors, the findings of each case study background are described below.

Table 1: Company characteristics

| Malaysian Halal LSPs | Case A | Case B | Case C |
|---|---|---|---|
| **Context of operation** | | | |
| Halal transportation activities | The services offered include Halal transportation, Halal distribution, Halal shipping, Halal freighting for sea and air cargo, samak service for containers, customs facilities and other Halal value-added services. | IT system has been enhanced with the critical check point list and traceability functions at the receiving process (verification of Halal status of cargo and labeling). No specific Halal transportation and distribution services yet. | Halal product handling - Halal product. Non-halal product at different space allocation. |
| Outsourcing Halal services | Outsource other 3PL for managing non-Halal item. | Samak services to other samak's contractor | Samak services to other company |
| **Technology and co-ordination** | | | |
| Existing ICT used to monitor Halal controls | RFID, bar coding, Internet real time tracking and tracing using GPS, TMS,WMS, EDI, CCTV | Bar-coding, Transportation Management System (TMS), Warehouse Management System (WMS), EDI | Bar-coding, Transportation Management System (TMS), Warehouse Management System (WMS), EDI |

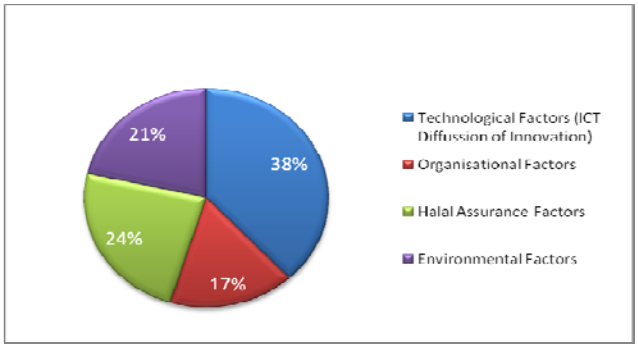



Figure 2: Influencing Factors in ICT adoption Process among Malaysian Halal LSPs

According to the framework of Halal transportation technology adoption (Figure 1), we evaluate the assent degree by interviewer on the adoption and utilization of ICT that complies with Halal standard in their organizations. We interviewed three Halal LSPs which have technology adoption experience for monitoring Halal in transportation and logistics. We score on the adopting factors under this framework according to factors that more explicitly expressed by interviewer in the case. Figure 2 shows the result in percent while Table 2 demonstrates the result of the study in brief.

Table 2: Influencing factors in the study

| Dimension | Factors found in the study | CsA | CsB | CsC |
|---|---|---|---|---|
| **Technological Innovation Factor** | Relative Advantage | • | • | • |
| | Compatibility | • | • | • |
| | Complexity | | • | |
| | Cost | • | • | • |
| | Image | • | | |
| **Halal Assurance Related Factor** | Syariah Law | • | • | • |
| | Halal Assurance Integrity | • | • | • |
| | Positive Intention Towards Halal Consumption | • | • | |
| **Organizational Related Factor** | Awareness - willing to adopt | • | • | • |
| | Organization Size | • | | • |
| | Organization Readiness | • | | • |
| **Environmental Related Factors** | Halal Logistic Standard on Transportation MS2400 | • | • | |
| | Relative Incentive offered for Halal Business | • | • | • |
| | Halal Program Initiatives | • | • | |
| | Environment Uncertainty | | • | • |

*Note:* • shows that the factor is explicitly expressed in the case.

## Technological Related Factors

The usage of the existing ICT that the LSPs have used to monitor Halal control in transportation's operation can be categorized into three areas: data communication technologies, identification technologies and, data acquisition technologies. All Halal LSPs have chosen data communication technologies and identification technologies as an ICT application for monitoring Halal. On the other hand, only CsA has applied data acquisition technologies that are CCTV to keep track the movement of product in their warehouse. Besides, all cases also have used the Internet based service platforms to extend their Halal market as a core business activity.

Relative advantage: Depending on the goal and the capability of LSPs, the ICT choice and its usage are depending on the compatibility of ICT characteristic with

the Halal requirement and guideline. For example, the RFID characteristic is tracking and tracing are seen as compatible with the Halal transportation guideline- track and trace goods along the supply chain. Besides that, the study regards the expectation of benefits of new ICT that Halal LSPs try to adopt as the perceived benefit. Many Halal LSPs (CsA, CsB, and CsC) have known the benefit of ICT adoption through previous experience – customer are more confident when they see some ICT investment applied as value added service to their product handling. From the result, all the Halal LSPs agree that relative advantage of ICT is the main contributing factor to adopt ICT.

Compatibility: All three cases applied management system that helps them to interact with immediate supplier or customers. For example, CsA said that an issue about container is very crucial. *"We wouldn't know what types of products that the container has carried because the container travels around the world. It could be anything that is non-Halal "*. Besides, all Halal LSPs (CsA,CsB,CsC) does not invest new technology just to cater Halal requirement. They use existing technology whereby the characteristic of these technologies is compatible and suitable for Halal requirement. For example, the characteristic of RFID is tracking and tracing. As stated in Halal logistic standard on transportation, the traceability system applied shall enables the identification of goods identification of the inbound goods and/or cargo for the processing stages from the immediate suppliers and the distribution routes at destination of the goods and/or cargo. This requirement is suitable with RFID characteristic. CsA, CsB addressed the main criteria of Halal controls in transportation are the information of location of goods being transported along the supply chain can be traced and the identification of product (information and specifics details). On the other hand, CsC explains system that suitable for product handling in ports is logistic system that based on segregation instead of detection. CsC also adds that ICT that can access and deliver the information faster are compliments. For example the information of goods and/or cargo (route) in the transportation chain services can be accessed efficiently.

Cost: Even in the Halal LSPs already using existing ICT for Halal control, cost is still a critical barrier. Even the adopter of ICT (CsB and CsC) is unwilling to upgrade the information systems or to adopt other advanced ICT service applications because of the high adoption cost. Due to limitations of capability or time, (CsB) outsourced a part of their business to private consultant. However, CsA explains that the cost for using ICT does not seem to be a barrier because they believe that it is a kind of investment. This thought seems to come from the higher awareness of improvement services after they achieved the benefit and the satisfaction of the ICT usage.

Complexity: The technological together with Halal knowledge of employees and their management capability can be a barrier to the adoption and extension of the information systems. Some CEOs (CsC and CsB) are worried about the introduction of new advanced ICT (just to cater Halal controls) because of the fear that their employees might be not familiar with it. Nonetheless, this factor does not seems to be a critical factor to adopt new ICT because they can ask the ICT service providers what they want and request an expert to train their employees.

No real time Halal tracking and tracing: Some Halal LSPs (CsB and CsC) addressed that all the monitoring Halal controls is still in manual. There is no real time Halal tracking and tracing. However all Halal LSPs (CsA, CsB and CsC) apply existing ICT to cater tracking and tracing issues.

## Halal Assurance Related Factors

These factors can be critical factors to directly or indirectly adopting and extending implementation of new ICT service applications for Halal. These factors are actually differentiating the study from other LSP adoption ICT studies. CsA and CsB agreed the changing business model to Halal business is because of their responsibility and the positive intention (towards Halal) as Muslims, providing Halal assurance in logistics services to customers from farm to fork. Manager of Halal in CsA answered, '*We believe that there are many customers in Halal market. Nowadays peoples are starting to be concerned with Halal issue and the demand for Halal is also increasing. Therefore, it is better for us to start now. It is our responsibility as a Syariah compliant company*'. However, CsC believed that the concerned of Halal or non-Halal issues is a small matter but the way companies handle the product in a good system and follow the Halal standard are the vital issues.

According to Tierman [10], for an effective logistics management of a Halal supply chain it is important to have Halal assurance system into the logistics strategy. Tierman said that, the company should have a solid visibility of its supply chain, supported by key performance indicators and finally, regular Halal logistics audits should take place to ensure that the Halal logistics performance is under control. All Halal LSPs (CsA, CsB and CsC) bear with this Halal integrated strategy therefore trying to achieve it with ICT assistance and implementation especially to cater visibility of its Halal

supply chain. They realized that information visibility is crucial between suppliers and within the company also. This is particularly important as poor performance on Halal has shown to have major implications for the image of the company and its brand, which can take years to recover. Halal logistics and Halal transportation guidelines are therefore important to be addressed in the contract between shipper and logistics service provider.

## Organizational Related Factors

Awareness – willing to adopt: This study estimates the awareness by using the intention of ICT adoption and of Halal business extension via ICT. Therefore, it is possible to assume that the Halal LSPs (CsA, CsB, and CsC) in the study have the positive awareness of ICT adoption in the light of their intention and efforts to adopt ICT for Halal monitoring. They also invest in technology, which plays an important role in providing product traceability through the storage of data such as product designated code, batch manufacturing number, expiry date, etc. The result shows that higher Halal awareness and awareness about ICT has played a very important role in ICT adoption and extension.

Organizational Readiness: This factor is the existence of external and internal information system that provides the development of customer and partner relationship management mechanism for Halal. CsA and CsC also felt IT maturity and organizations readiness is the major affecting factor when planning to adopt an ICT.

## Environmental Related Factors

Government Incentives and Support

a) Halal logistic standard on transportation MS2400 (2010):

All Halal LSPs in the study are Malaysia's Halal Jakim-certified logistics provider and also apply for Halal logistic standard that covers on transportation, warehouse and retailing. This is a driving factor as to comply with the standard; the organization must establish and apply a traceability system that enables the identification of goods and/or cargo in the transportation chain services, the identification of the inbound goods and/or cargo for the processing stages from the immediate suppliers and distribution routes at destination of the goods and/or cargo. [20]. Some Halal LSPs (CsA, CsB) already apply the traceability system for fulfilling the standard requirement. This can be seen as influencing factors that impact the process of ICT adoption.

b) Relative Incentive offered for Halal Business and ICT application

In this study, there is not enough mention about the government support toward ICT application. However, most Halal LSPs desire various and appropriate support from the government. Some of Halal LSPs (CsA and CsC) have used one of the ICT service platforms developed by the support of the government According to Malayisan Logistic Directory [5], in terms of government support, attractive tax incentives are offered for businesses involved in Halal products and services. Under the incentives, Halal logistics operators are eligible for: (i) Full income tax exemption for a period of five years; OR 100% income tax exemption on qualifying capital expenditure for a period of 5 years. AND (ii) Exemption on import duty and sales tax on equipment, components and machinery used directly in the Cold Room operations subject to current policies. Realizing these benefits, most Halal LSPs (CsA, CsB and CsC) take these incentives as to encourage new investments in their Halal business. Besides, Halal LSPs also gains some benefit if use modern and state-of-the-art machinery, ICT and equipment in producing high quality 'Halal' in their logistic activities that comply with the Halal standard. Halal LSPs (CsA) spending RM7 million to RM10 million in 2011 as part of its expansion plan to meet rising Halal logistics demand nationwide. So, positive government supports and roles are among factors that may impact the ICT adoption process.

c) Halal Program Initiatives

All Halal LSPs (CsA, CsB and CsC) in the study have used an ICT service platform provided by Halal Development Centre (HDC), which have played an important role in providing many programs and training related to ICT adoption in Halal sector. The training helped Halal LSPs on how to comply with the Halal logistic standard. They were organized in 2006 by the enforcement of government legislation to promote ICT adoption among LSPs. To continuously ensure Halal integrity in its service, (CsA, CsB, CsC) has increased efforts and resources in facility maintenance and training for its employees. This initiative is to create awareness and knowledge amongst its employees in managing halal products as well as ensure its effectiveness in terms of application. The CsB and CsA will send their Halal officer to join the Halal training course once a year to master in Halal transportation's issues for example sharing container, lack of visibility across an entire supply chain

includes poor container identification, segregating allocation between Halal and non-Halal product in same container (for contamination avoidance).

Environment Uncertainty: As local Halal food industry was worth RM45 billion while the global Halal market was valued at about RM2 trillion, the competition between Halal LSPs leading in Halal business are some of the contributing factors. During the interview CsB stated that some Halal LSPs (CsA, CsC) are one of their competitor are moving forward in applying modern ICT to monitor their Halal transportations operation and logistics activities. They (CsB) see this as a new challenge for them to beat their competitors.

## 5. Discussion

The Halal LSPs in Malaysia have experienced various benefits from the usage of the ICT in their Halal logistic activities. There are two main reasons why the Malaysian Halal LSP's wants to adopt ICT in monitoring Halal integrity; first, to assist assurance of Halal integrity throughout the supply chain, and second, to increase efficiency of the logistics performance. These are very similar to the benefits found in other literature which include the increase in Halal transparency during food production, increase in consumer trust on the Halalness due to the increase in the amount of information about the production process, food-safety control [26], and making information available along the supply chain [13]. Better organization of supply chains increased the Halal performance at the destination [10] and ease of access to know Halal status in a few second without cost. Moreover, the adoption of ICT among the Halal LSPs is positive if the characteristic of ICT are compatible with the Halal requirement.

Among technology related factors, the cost for using ICT currently in most Halal LSPs think that it is a kind of investment. This thought seems to come from the higher awareness of improvement services after they achieved the benefit and the satisfaction of the ICT usage. In addition, they are willing to adopt other ICT to extend the market and to pursue the efficacy of their Halal business. In this case, however, cost is still a crucial barrier to adopt ICT even in the LSPs that are already adopting the ICT service platforms and where their awareness and the intention of ICT adoption are increasing. It also shows that the cost of ICT adoption could be decreased gradually, depending on the level of ICT development and the degree of the assistance of the external environment.

Of environmental related factors, it can be seen that the more the provision of government support, the more positively that Malaysian Halal LSP will adopt innovation in Halal logistic technology. The relative incentives from government encourage Halal LSPs to invest in Halal market. Also for organizational factors, the higher the awareness of the top managements and their readiness to adopt ICT in their organization will influence the ICT diffusion in Halal operation.

As an indirect factor impacting ICT adoption in the Halal LSPs, all Halal related factors shows the motivation of Halal LSPs to adopt ICT for monitoring Halal. In the analysis of data obtained from interviews, three technological components can be identified that suit and compatible for Halal transportation:

a) Location tracking – system to determine location of Halal goods being delivered to customer
b) Identification of product – system to identify the information of goods (i.e history from immediate supplier), which in the basic form, automatic identification technologies help to collect the shipment identification number and information, and provide this information as an input to the rest of the system.
c) Data Communication - technologies to access and deliver the information

In the case of Halal transportation's operation (HTO), information could mean the Halal goods location, point of origin and destination, the content, the inspection results, etc. These traceability systems can be used to provide real time global Halal information for internal use in terminal operation or in transportations chains. For instance client may want to have accurate Halal information on where their containers are and when they arrive, and the governments may desire to ensure that cargo arriving on land is properly taxed or dangerous goods are not smuggled.

Based on the above discussion and the research framework as shown in Figure 2 and Table 2, we propose the propositions as follow:

**Proposition 1:**
The implementation of HTO is positively related to Halal transportation standard (MS2400l:2010)

**Proposition 2:**
The more explicit the technology towards HTO requirement, the more likely that Malaysian Halal LSP will adopt innovation in Halal logistic technology / ICT

**Proposition 3:**
The more the organizational encouragement, the more likely that Malaysian Halal LSP will adopt innovation in Halal logistic technology / ICT

**Proposition 4:**
The more the environment uncertainty, the more likely that Malaysian Halal LSP will adopt innovation in Halal logistic technology /ICT

**Proposition 5:**
The more the provision of government support, the more likely that Malaysian Halal LSP will adopt innovation in Halal logistic technology / ICT

## 6. Conclusion

This research analyzed the collected qualitative data, given the exploratory nature of the study. The research outcomes show that technology related factors and Halal related factors contributed positively to efficient Halal LSPs operation while the government factors are very important in the reduction of the main barriers and the creation of the atmosphere of ICT adoption in Halal LSP sector. These factors are the influencing factors that give impact to the process of ICT adoption in Halal transportation. In addition, this study also categorized technological components of ICT adoption for Halal controls in transportation into three types: location and tracking, identification of goods and/or cargo and data communication.

This study proposes guidelines for logistics service innovations in the area of logistics and Halal transportations. Besides, the study discussed the improvement in Halal services after they achieved the benefit and the satisfaction of the ICT usage which have taken into consideration several factors. This study also contributes to the advancement of knowledge through the application of Halal concept into logistics service practices. The needs to be innovative in initiating more logistics services that are based on Halal concept are crucial in meeting the needs of the increasing demand by the customers especially the Muslims. The findings provide insights to the practitioners of the importance to be innovative in maintaining Halal integrity along the supply chains to fulfill the growing demand of the Halal products.

## References

[1]. Tieman, M., *The Future of Halal Logistics Solutions*, in *The Halal Journal*. 2006, KasehDia Sdn Bhd.

[2]. Halal Development Corporation (HDC). *Support Infrastructure - Halal Logistics*. 2009 [cited 2009 4 Aug]; Available from: http://www.hdcglobal.com/portal/mainpage.php?module=Maklumat&kategori=49&id=242&papar=1&id2=4&menu=168.

[3]. Jaafar, et al. *Innovation in logistics services (halal logistic)*. in *Proceedings of the 16th International Symposium on Logistics (ISL), Berlin, Germany (2011):* . 2011.

[4]. Tierman, M., *The Building Blocks of A Halal Transportation System.* The Halal Journal, 2009.

[5]. Malaysia, D.o.S. (2011) *Malaysia Logistics Directory 2011/2012*. Halal Logistic, 9.

[6]. Rashid, M.A.a.A.-Q., N. A. (2001). . , , *E.Commerce Technology Adoption Framework by New Zealand Small to Medium Enterprises.* Research Letters Information Mathematical Science, 2001. **2**(1): p. 63-70.

[7]. Lee, S.W. and D.J. Kim. *Driving Factors and Barriers of Information and Communication Technology for e-business In SME's: A Case Study in Korea*. in *IADIS International Conference e-Society* 2004.

[8]. Susana Garrido Azevedo, João Ferreira, and João Leitão, *The Role of Logistics Information and Communication Technologies In Promoting Competitive Advantages of The Firm.* 2007.

[9]. Mazlan, M., *Innovation Diffusion and ICT Adoption in Jakim Halal Certified Company in Klang Valle*, in *Faculty of Information Technology And Quantitative Scienc*. 2006, University Technology Mara (UiTM): Shah Alam, Selangor. p. 62.

[10]. Tierman, M. (2010) *Halal Logistics -Logistics Insight Asia, 1/1/2010*. Logistics Insight Asia.

[11]. Halalan-Toyyiban, F.K.K.P., *HALALAN TOYYIBAN ASSURANCE PIPELINE – Management System Requirements for Transportation of Goods and Cargo Chain Services*, S.M. (SM), Y.E.S. (YES), and H.I.D.C. (HDC), Editors. 2010.

[12]. Tierman, M., *Halal Transportation - The building blocks of a Halal transportation system* in *The Halal Journal - Jan/Feb 2009*. 2008, The Halal Journal.

[13]. Husny, Z.J.M., *The Needs of Halal Transportation Control in Malaysia: A multiple case study approach.*, in *Faculty of Built Environment.* 2010, Universiti Teknologi Malaysia: Skudai, Johor. p. 137.

[14]. Lin, C.-Y. and Y.-H. Ho, *Technological Innovation for China's Logistics Industry.* Journal of Technology Management Innovation, 2007. **2**(4).

[15]. Lai, et al., *Information Technology Adoption in Hong Kong's Logistic Industry.* Transportation Journal, 2005. **44**(4): p. 1-10.

[16]. Tang, L.-L. and W.-C. Tsai, *RFID adoption Model for Taiwan's Logistic Service Providers.* 2009.

[17]. King, N. and N.R. Anderson, *Innovation and change in organizations.* 1995: London: Routledge.

[18]. Damanpour, F., *Organizational innovation: a meta-analysis of effects of determinants and moderators.* Academy of Management Journal, 1991. **34**(3): p. 555-590.

[19]. Galliers, R. D. & Sutherland, and A. R., *Information Systems Management and Strategy Formulation: Applying and Extending The Stages of Growth'*

*Concept in Strategic Information Management: Challenges and Strategies in Managing Information*, ed. G. 2nd. ed. (Eds, R.D., Leidner, D. E. & Baker, B. S. H.) Butterworth-Heinemann, Oxford. 1999.

[20]. Standard, M. and YES, *Draft Malaysian Standard - HALALAN TOYYIBAN ASSURANCE PIPELINE – Management System Requirements for Transportation of Goods and Cargo Chain Services*. 2010.

[21]. SIRIM and Berhad, *Standard and Quality News*, in *Standardisation of Halal food*, SIRIM and Berhad, Editors. 2004.

[22]. Norman Azah Anir, Md Nasir Mohd Hairul Nizam, and A. Masliyana. *RFID Tag for Halal Food Tracking in Malaysia: Users Perceptions and Opportunities* in *7th WSEAS Int. Conf. on TELECOMMUNICATIONS and INFORMATICS (TELE-INFO '08)*. 2008. Istanbul, Turkey.

[23]. Rogers, E.M., *Diffusions of Innovations.* 4th. Ed. New York:, 1995.

[24]. Yin and R. K., *Case study research: Design and methods.* , ed. C.S. Thousand Oaks. 1994.

[25]. Garry W. Auld, et al., *Development of a Decision Tree to Determine Appropriateness of NVivo in Analyzing Qualitative Data Sets.* Journal of Nutrition Education and Behavior, 2007. **39** (1): p. 37-47.

[26]. Zailani, S., et al., *Halal Traceabiliy and Halal Tracking System in Strengthening Halal food Supply Chain for Food Industry in Malaysia (A Review).* Journal of Food Technology, 2010. **8**(3): p. 74-81.

**First Author** Mohd Iskandar Illyas Tan . Received his Master Degree in Computer Science in 2003. Lecturer at Department of Information Systems, Faculty of computer science and information systems. Head of HOLISTICS Lab. Currently working on his Phd in Information Systems and active in doing research in Halal and Logistics.

**Second Author** Raziah Noor Razali holds a Bachelor's degree in Computer Science from Universiti Teknologi Malaysia Skudai, Johor. Prior to further their studies to degree level, she holds a Diploma in Computer Science (Multimedia) from Universiti Teknologi Malaysia Kuala Lumpur City Campus. She was a programmer and systems analyst as a part time job in private companies. Currently, she is in the final year Masters in Computer Science at the Universiti Teknologi Malaysia Skudai, Johor. Her research areas are Halal logistic technologies and Halal transportation technologies.

**Third Author** Mohammad Ishak Bin Desa is a Professor at Department of Modeling and Industrial Computing (PPI), Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. Received his PhD in Operational Research from University of Salford . He is currently the head of Operations and Business Intelligence Research Group (OBI). Since 2010, he has been a member of HOLLISTIC research group.

# Dynamic Replica Control Algorithm for Periodic/Aperiodic Transactions in Distributed Real-Time Databases

**Torky Sultan[1], Hazem El bakry [2] and Hala AbdelHamed[3]**

**[1] Information System Department, Helwan University, Cairo Egypt**

**[2] Information System Department, El-Mansora University, El-mansora, Egypt**

**[3] Information System Department, El-Mansora University,Cairo, Egypt**

**Abstract**

Maintaining consistency between the actual state of the real-time object of the external environment and its images as reflected by all its replicas distributed over multiple nodes is one of the most important issue affecting the design of real-time database. Efficient replica control algorithm can contribute significantly to maintain this consistency. A replication control algorithm is presented for medium and large scale distributed database system by trying to maintain an independent consistency degree for each data object by presenting an adaptive dynamic replication control algorithm. It is based on some system factors to increase the chance of having an updated data item locally; avoiding remote access to meet timing constrains and achieves both availability and consistency for the replicated data as much as possible. A detailed simulation study shows that our algorithms can greatly improve the system performance compared to the systems either without replication or with full replication.

***Keywords:*** *Distributed database, Real-time databases, Real-time transaction, and Replica control algorithm.*

## 1. Introduction

Many real-time systems are inherently distributed in nature, and need to share data that are distributed among different sites. For example, military tracking, medical monitoring, naval combat control systems and factory automation etc. All of those critical systems need data to be obtained and updated in a timely fashion [1]. But, sometimes data that is required at a particular location is not available when it is needed and getting it from remote site may take too long before which the data may become invalid. This potentially leads to large number of tardy transactions (transaction that miss their deadline).

One of the solutions, for the above-mentioned problem, is *replication* of data in real-time databases. By replicating temporal data items, instead of asking for remote data access requests, transactions that need to read remote data can now access the locally available copies. This helps transactions meet their time and data freshness requirements. In order to suite the different needs of the distributed real-time systems such as different data workloads and database specifications, multiple ways to handle the replication control and different replication schemes are proposed.

In the distributed systems, replication is seen as a cost effective way to increase availability. However, replication is used for both performance and fault-tolerant purposes thereby introducing a constant trade-off between consistency and efficiency. There are two main approaches for replication; *synchronize* (also called Active or state machine) in which a collection of identical servers maintain the same copies of the system state, client write operations are applied synchronously to all of the replicas [2,3,4]. Although this approach increases consistency of the replicated data, it increases system overhead. *Asynchronous* (also called lazy or passive) replication on the other hand where changes introduced by a transaction are propagated to other sites only after the transaction has been committed. However, this approach reduces system overhead at the expense of temporal consistency.

Replication algorithms can also be characterized according to *what* and *where* the objects are replicated. The most extreme is *full* replication in which all data items are replicated to all sites in the distributed system. The benefit of full replication is that all data are available to read locally, thus, leads to increasing performance. But this slow down the system since updating one copy creates transactions for updating all other sites, also issues like concurrency control and recovery become more

complicated. Regarding the locations and number of replicas, we can distinguish between *static* replication algorithms in which both locations and number of replicas are fixed and *dynamic* replication when the locations and number of replicas are dynamically changing according to system conditions and data needs.

In that work, a replication control algorithm for medium and large scale distributed database system is presented, by trying to maintain an independent consistency degree for each data object by presenting an adaptive dynamic replication control algorithm. This algorithm is based on the entire system workload of the distributed site and increase the chance to have an updated data item locally to avoid remote access to meet timing constrains and achieves both availability and consistency for the replicated data as much as possible.

## 2. Related Work

In replicated database systems, copies of the data items can be stored at multiple sites, which achieve two complementary features: performance improvement and high availability. On the other hand, data replication introduces its own problems; Access to a data *item* is no longer controlled exclusively by a single site, instead the access control is distributed across each site storing a copy of the data item. It is also necessary, to ensure that mutual consistency of the replicated data is provided, in other words, replicated copies must behave like a single copy. This is possible by preventing conflicting accesses on the different copies of the same data item, and by making sure that all data sites eventually receive all updates. Therefore, major issue is the development of replication protocol/policy. The problem of finding an optimal replication scheme in a general network (i.e., a replication scheme that has a minimum cost for a given read-write pattern), has been shown to be NP-complete for the static case.

Several classifications are possible for replication [4, 5, 6, 7, and 8]. In [4] Wiesmann & Schiper distinguish five different techniques (active, weak-voting, certification-based, primary copy and lazy replication). All these techniques dealt only with fully replicated databases and need a reliable total order broadcast in order to propagate the transaction updates.

All the replication models that have been developed so far can be classified into *optimistic* replication or *pessimistic* ones [9]; in the optimistic replication, all the operations are performed locally at each node as if there were alone in the system and optimistically assumes that there is no conflict with the other replicated copes located in the other nodes. This approach will increase both response time and performance by avoiding all the remote access. But, it leads to temporal inconsistency between different nodes which require a good conflict resolution and compensation polices to eliminate this inconsistency. Other models are pessimistically avoiding conflict between concurrent operations running in the other nodes by different methods such as, global lock or primary copy in which only the primary site is permitted to update its items.

There have been a number of research papers about data replication in traditional database systems where some sporadic efforts have been made for the development of different types of protocols/policies [10, 11, 12, 13, 14], but it is not for real-time systems. In the literature there is a little replication models for real-time database systems [15,16,17], one of those models [ORDER]is found in [18] where full replication is used in medium-scale or large-scale distributed real-time database systems. It presents the ORDER algorithm that is designed to work in an environment where all data types and relations in the system are known a priori. In term of scalability, the algorithm has been enhanced to a replication algorithm called On-demand Real-time Decentralized Replication with Replica Sharing (ORDER-RS). In [19], Peddi *et al.* present a replication algorithm called Just-In-Time Real-Time Replication (JITRTR), which creates replication transactions based on client's data requirements in a distributed real-time object-oriented database. In [20] a replication protocol named PRiDe designed for optimistic replication with forward conflict resolution in distributed real-time databases was described. The model defines four phases for the replication protocol: the local update phase, the propagation phase, the integration phase, and the stabilization and compensation phase. In that model, the transactions are executed locally and the replicated items are updated as if they are alone in the system, after completing the transaction, the model check s for conflict in the other nodes and a conflict resolution policy is used to resolve it.

## 3. Model Overview

In this paper, a replica control algorithm used in DoMORE (Dynamic allocation Module for Replication in Distributed Real-time Database) is described. DoMORE is a replication model based on increasing the probability of providing the updated data for the real-time transactions to meet their timing constrains by increasing their chance to have these data locally without the need to get it remotely from other sites. This goal can be achieved by dynamically allocating data to distributed nodes according to their access pattern. The model also allows different degree of consistency for each data object which is dynamically

calculated according to different factors (entire workload and system requirements). DoMORE model allows most of the transactions to execute and commit locally as if the transaction is executed in a local, centralized database. It uses a strict-2PL commit protocol for distributed transaction, upholding the ACID properties [21], and transaction processing is guaranteed to be serializable with respect to other local transactions.
As it was mentioned, the objective of the replication model is to give the clients the illusion of service that is provided by one server and the clients have no knowledge about the data existence behind. Of course, maintaining redundant data adds overhead to the system, and this can be reduced by exploring week consistency semantics of applications. This tradeoff between consistency and system cost is the main problem of all the replication models.

Generally, consistency is a term for discussing the correctness of data in a database, the database is said to be consistence if all consistency predicates are hold. For real-time database, consistency predicates can refer to the relationships between database objects and external environment that are modeled by the database from time point of view. In a replicated system we can define three different types of predicates for consistency[22]; *external* temporal consistency which deals with the relationship between an object of the external world and its image on the database, *inter-object* temporal consistency which is the relationship between different objects or events (within a single node), and it also includes the relationship between temporal data item and non-temporal data item that depend on that item, and *mutual* consistency, which reflects the relationship between the object and its copy (replica) in different remote sites.

DoMORE employs both eager and lazy replication according to the types of database items, and it guarantees that all the transactions will read an updated valid data items and maintain both Temporal Consistency and Mutual Global Consistency [23, 24, 25]. In DoMORE, global consistency is achieved through continuously propagation and integration of updates (typically, transaction updates are propagated and integrated as soon as possible, but propagation or integration may be deferred under certain circumstances, such as if there is a transient overload).

DoMORE is also based on the concept of Virtual Full Replication (ViFuR) [26] that has been introduced in DeeDS [27, 28]. It creates a perception of full replication by ensuring that all used data objects are available at the local node so that user can interact with the database as if it is full replicated and all data are available locally. Thus this approach can take the advantages of full replication such as, reducing the resource usage compared to full replication, and transaction timeliness, simplified addressing of communication between nodes, as well as support for fault tolerance. Accordingly, the database user cannot distinguish a virtually fully replicated database from a fully replicated one.

### 3.1 System Model

The system model consists of a group of distributed main memory real-time databases connected by high-speed networks. It was assumed that a reliable real-time communication is maintained, i.e., any messages sent over the network is eventually delivered and have predictable transmission time [25]. The whole database is segmented on different nodes, we can define segment as a group of data objects that share properties, capturing some aspects of the application semantics, and is allocated to a specified subset of the nodes (possibly temporarily inconsistent with each other). Each segment is considered as a partition of the database and as units of allocation of replicas, which can be individually replicated based on specified replication requirements from all the clients at a certain database node. If the specification indicates that a data object will never be used by any clients on a node, it does not need to be replicated to that node, and also a certain database object may not be available at a node, but the clients at the node do not need to be aware of that, because they will never access it. This is called call virtual full replication where the client assumption of full replication of the database is still valid.

Traditional RDBMs are based on the assumption that data resides primarily on disk and in a dynamic runtime environment data might be on disk or cached in main-memory at any given moment. Because disk input/output (I/O) is far more expensive than memory access, main memory databases have been used because of the high performance of memory accesses and the decreasing cost of main memory [29,30]. Because access to main memory is so much faster than disk access, we can expect transactions to complete more quickly in a main memory system. So, in the distributed transactions that use lock-based commit protocol locks will not be held as long, thus, lock contention may not be as important as it is when the data is disk resident.

In the system, a firm real-time database model is used; tardy transactions (transactions that have missed their deadlines) are aborted. Fig 1 illustrates the main components of the model; the monitor is periodically collects the workload data of the entire system and sends them to the admission controller. It also, checks the system performance periodically and records the statistics data about the transactions' miss ratio. The admission controller is responsible for accepting or rejecting the remote requests from the other nodes. Transaction manager is

responsible for generating local update transactions and replication transactions for their nodes. We assumed here that transactions are executed sequentially and there is no concurrent transactions, however, if concurrent transaction execution is required, the algorithm can be extended to allow concurrent transaction execution, e.g., through a locking scheme. In the next work we will consider concurrent execution of transactions and use one of the concurrency control protocol suitable for real-time systems. In the priority assignment and scheduler component, it is known that scheduling is necessary to choose an action to execute when several guards are enabled. So, it is necessary that each transaction is assigned a priority (such as prioritizing local operation), local read and update actions should have precedence over propagation and replication actions. For simplicity, in this paper, Earliest Deadline First (EDF) policy is used and transactions are processed accordingly.

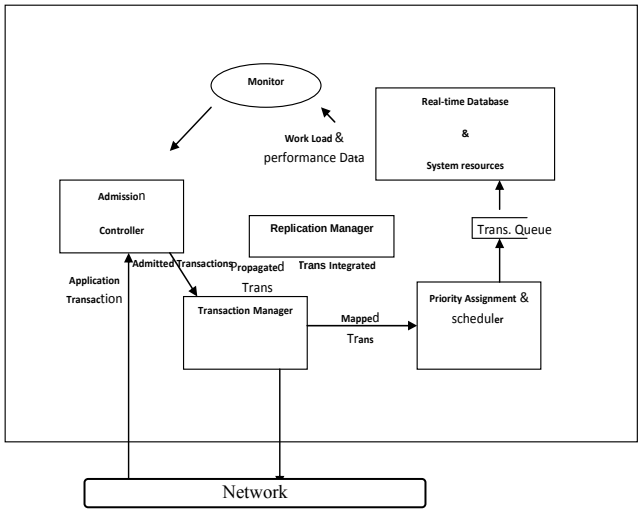



Fig 1: The inter model components at node N

As mentioned above, the model implements both eager and lazy replication, for eager replication where a synchronies replication is performed, all the sites participated in a transaction's execution are engaged in an atomic commit protocol (ACP). The model use a strict 2PL committing protocol to ensure consistent termination of distributed transactions despite site and communication failures. The two-phase commit (2PC) protocol [31] is the simplest and most used ACP. The primary goal of a two-phase commit protocol is to ensure that all participants agree on whether a transaction commits or aborts. Since 2PC consumes a substantial amount of a transaction's execution time due to the cost of its coordination messages and forced log writes to stable storage required for recovery, a number of 2PC variants appear in the literature, most notably, presumed abort (PrA) and presumed commit (PrC) [32]. As opposed to PrA, PrC has been designed to reduce the cost associated with committing transactions rather than aborting ones.

## 3.2 Data Model

The data model used in this paper categorizes the data used into two main categories; *Sensor* Data objects, as the real-time systems interacts with the environment through various sensors, e.g. temperature and pressure sensors and it is important to maintain consistency between the states of the environment as perceived by the sensor and with the actual state of the environment. The sensed data is processed further to derive new data called *Derived* Data that depends on past sensor data, for example the temperature and pressure information pertaining to a reaction may be used to derive the rate at which the reaction appears to be progressed which is in turn could be used to derive a new data. So we can define two different types of data items; Temporal data items that changes with time and have a validity interval and Non-temporal data items which is not change with time and thus they don't have a validity interval. If we define each site or node as a segment for a set of data items, it is called a primary site for them (Psite). For a specific data item, the copy of data item at the primary site is called *primary* copy and the copies that are replicated are called *replicated* copies or *replicas*. As illustrated in a previous work [33], the proposed database framework will be used, whereas there are two types of data located at each node; local or shared data. Local data object can only be updated by its primary site, and the shard data items can be updated by any site in the network in cooperation with its primary site. It was assumed that for each site, a backup site is predefined to be used as a backup site in case of site failure and to guarantee the minimum degree of replication.

Each replica has associated a version number (VN) which reflects the last update number for this data object. When an update is received, the receiving node (primary) increases the updating node's Version Number of local replica of the updated object. Any object has the following specifications shown in figure2, each $o \in O \exists o = (Id, Type, Name, PsiteId, Value, TS, VI, BUF, VN, FR)$

| Id / type / name/ PsiteId / VI / BUF/ FR | | |
|---|---|---|
| Value<br>Current Value | TS<br>(Time Stamp | VI<br>Validity Interval |

Fig 2: the structure of real-time attribute

*Id*: is a unique identifier for the object on his primary site.
*Type*: whether it is local or shard data object.
*PsiteId*: the object's primary site id where the object was originates, this attribute gives an indication of whether the object is a primary data object or it is a replica e.g., if PsiteId = local site, this object is a primary object

originated at this site, otherwise it is a replica for a remote data object.
*Value*: is used to store the final attribute value captured by the related last update method.
*TS*: is used to store the last time at which the attribute's value was updated.
*VI*: denotes object's absolute validity interval i.e., the length of the time interval following timestamp during which the object is considered to have absolute validity.
*BUF*: is the Basic Update Frequency, for each temporal data object it is updated periodically at a given update frequency received from its primary.
*FR*: A predefined freshness requirement to maintain the consistency level between different replicas scattered over all sites for the same data object.
*VN*: is the version number that reflects the last update number of that object.

### 3.3 Transaction Model

For the local type data objects, the Read Phase is performed locally at each site for only the active replicas located in this site. The Propagation phase of the replication process for any transaction updates of a local data item to remote nodes is delayed until after the transaction commits. The propagation messages for remote transactions that have been received at a node are integrated locally according to a local scheduling policy.

For the shared data item the commitment of the transaction that update it is conditional by at least the agreement of its primary site to which it is belongs using the (2PC) protocol. In that case, the integration task maintains local conflict detection data structures and is responsible for making updates by remote transactions visible to local transactions. The integration task is serialized with respect to local transactions. A transaction T and all of its updates are said to be integrated on node N if T has been committed locally and propagated to N from the other node and has been processed by the local integration task on N. The transactions are divided into read (query) transaction in which all its operations are read the data objects, while the update transaction can contain at least one write operation. Transaction can be also classified into remote or local transaction; the transaction is considered local if all its operations are performed in the local site, and it is remote if at least one remote operation. Note that only transactions of one operation are considered here.

### 3.4 Formal Definition

Before we describe the algorithm, we need to define some terms formally used to describe the algorithm. When a temporal data item (whether it is local or shared data item) is updated in a specific node, the *Replication Degree* ($\mathcal{RD}$) defines the number of nodes to which it must be replicated and the number of propagation messages that must be created by the Replication Manager. *Replica Allocation Set* ($\mathcal{RAS}$) defines a set of sites or nodes to which the replica updates or the propagation messages must be sent.

*Definition* 1: Replication Degree $\mathcal{RD}$ is the number of sites/nodes to which the propagation messages will be sent for a particular update. It is calculated by a Replica Degree Function $\mathcal{RD}$F which takes specific parameters (Node Workload, Object Freshness requirements ($\mathcal{OFR}$), User Defined Level. Note that the upper bound for $\mathcal{RD}$ is the total number of nodes in the distributed system.

*Definition* 2: The *Replica Allocation Set (RAS)* of a propagation transaction $\mathcal{T}_{propagation} = (\mathcal{T}id, \mathcal{L}site\ Id, \mathcal{R}site\ id, \mathcal{WS}, \mathcal{GUF}, \mathcal{DL}, e)$ is the set of remote nodes hosting replicas of objects in the write set of $\mathcal{T}$. That is:

$$RAS(T) = \{n \in N | \exists o \in nT(R(o,n) \neq \varphi$$

To determine the $\mathcal{RAS}$, the model maintains for each data object -at its primary site- a new data structure called a *needlist*, which is an array that contains a list of *sites ID* requesting that data item, and is arranged by the highest frequency rated site for demanding that object.

*Definition* 3: Let $\mathcal{D} = (\mathcal{O},\mathcal{R},\mathcal{N})$ *be* a distributed replicated database, and let $r \in R$ be the replica of object $o \in O$ on node $n \in N$. The NeedList $\mathcal{NL}$ *(o)* for o is a vector of $|\mathcal{N}|$ elements containing the latest $\mathcal{N}$ site use this object. This vector is of the form $< \mathcal{N}_1id, \mathcal{N}_2id, ..., \mathcal{N}_{(N-1)}id >$, where each element $n, 0 \leq n < |\mathcal{N}|$ represents an identifier for a unique node or site use this data object recently.

```
void append(S_i (id)) // Append i^th Site to the end of
the needlist.
int HighestPriority () // Returns the SiteId located at
the head of the array.
void RAF (S_i (id)) // Append i^th Site to the RAS
void RemoveSite(S_i (id)) // Remove the i^th Site from
the needlist after performing the RAF function on it.
// HighestPriority () and RemoveSite () both return -1
if the queue is empty.
```

Fig 3 the methods performed in the need list

Needlist ($\mathcal{NL}$) implements the methods in figure 3, the first method for adding a new site id in the need list for the intended object, this method is implemented when the object is accessed or updated by that node. When adding a new site, it is added in the head of the array. The order of the elements located in the array indicates the priorities for selecting that site to be added in the $\mathcal{RAS}$.

*Definition* 4: For a propagation transaction $\mathcal{T}_{propagation}$ executing in site $n \in N$ the *Replica Allocation Function RAF:*

$n \rightarrow \mathcal{RAS}(\mathcal{T})$ is the function that maps a node N located in the head of the needlist to *Replica Allocation Set* of that transaction.

As illustrated earlier, each node N hosts a set of temporal data objects as a primary site, and also maintains a set of replicas of temporal data objects hosted by other nodes. All replicas of a particular data item are updated using the fresh value from their primary copy. When a replica existed in a remote site, and periodically receives an update from its primary site within its validity interval $\mathcal{VI}$ it is called an *Active Replica*, otherwise, it is called an *Inactive Replica*. An active replica will become inactive if it is not updated within its validity interval.

*Definition* 5: For a set of replicas $\mathcal{R}$ of logical objects in a set $O$, replica $r \in R$ of a logical object $o \in O$ (where o is a sensor data object) on a particular node is called *Active Replica* $\mathcal{R}(o,\mathcal{N})$, if :

$$(CurrentTime - TS(o)) \geq VI(o).$$

*Active Replicas* for the derived data objects are determined according to their relative consistency, for example if we considered two objects $O_1$ and $O_2$ which have two timestamps $\mathcal{TS}_1$ and $\mathcal{TS}_2$ respectively, $O_1$ and $O_2$ satisfied the relative consistency called *Relative Valid Interval* $\mathcal{RVI}$ if:

$$|TS_1 - TS_2| \leq RVI$$

*Definition* 6: For two replicas *r1, r2* where $ri \in R$ of logical objects $O_1$, $O_2$ in a set O, (where o is a derived data object) on a particular node is called *Active Replica* $R(O_i,N)$ if: $|TS_1 - TS_2| \leq RVI$.

## 3.5 Replica Control Algorithm

The goal of the proposed Replica Control Algorithm is to gain efficiency over Virtual Full Replication (ViFuR) strategy by dynamically changing the replication degree $(\mathcal{RD})$ and replica allocation set $(\mathcal{RAS})$. The main question of any replication model is *how to determine an appropriate replication level and placement for an object*? In some replication schemas the replication level for an object is predefined (e.g., 5 copies) leaving the run-time system to determine the placement of the five replicas in the network [34], while in others, it also specifies the locations of the replicas. These interfaces require the system designers to make a mapping from the desired characteristics of the (replicated) object, such as fixed level of availability, to a replication level and placement that will achieve those characteristics.

For using in this algorithm, a new replication schema is defined in which neither the replication degree nor the allocation sites is defined. Rather, for each object the replication degree and the allocation table is dynamically changing according to data access and system requirements at each site, e.g. if we have N nodes each has a set of data items (segment) that it considered as a primary site for them, the Replication Manager (RM) is responsible for dynamically calculating the replication degree $\mathcal{RD}$ that must be propagated to the other remote sites at each object update. $\mathcal{RD}$ =N-1 in case of full replication and $\mathcal{RD} \neq 0$ for fault tolerance purposes. $\mathcal{RD}$ is calculated by a separate module in the replication manager according to specific factors affecting this value (here, we consider only two factors; System workload and a predefined freshness requirement for each data object).

Using Work Load ($\mathcal{L}$) as a factor, the $\mathcal{RD}_L$ is calculated as follows:
If we have N sites to propagate a new replica, and we have 100 percentage to represent the entire workload for each node, we can divide that workload into n ranges, the difference between any two consecutive ranges is x, where x = 100/n.

$$100 - (RD_L * x) \leq L < 100 - ((RD_L - 1) * x) \quad 1 \leq RD_L \leq n \quad (1)$$

For example; if there are 5 sites in the network, and according to the entire workload, the range is calculated as follows: x = 100/5 where x=20% of workload. And when the entire workload is between 40% and 60%, then using (1) $\mathcal{RD}_L$= 3.

Because different factors can affect $\mathcal{RD}$ differently, the following weighted average equation can be used to calculate the value of $\mathcal{RD}$ to be used by the algorithm.

$$RD = ((W_1 * RD_L + W_2 * RD_{FR} + .. + (RD\ of\ m\ factors) * W_m))/m) \quad (2)$$

Where $m$ is the number of factors and $w$ is the weight for each factor and $(w_1+w_2.....+w_m) = 1$. The freshness requirement $(\mathcal{FR})$ for each object is taken as another factor affecting the Replication Degree (The $\mathcal{FR}$ is given for each object), accordingly, the $\mathcal{RD}$ can be calculated using (3).

$$RD = (W_1 * RD_L + W_2 * RD_{FR})/2) \quad (3)$$

The model divides the replication process into 4 phases, (Read Phase, Update Phase, Propagation Phase, and Cooperation and Integration Phase). As it was illustrated previously that data objects are classified to either local or shared data object. For the local data items, the primary site is responsible for both updating and propagating phases, while for shard data items, any site can update it, and only the primary site is taking the responsibility of the propagating phase.

When to update a replica, is another decision made by the model; the primary site begins pushing replicas to the

other sites when the primary site receives a new value for a specific data item from the external environment (sensor data). The algorithm calculates $\mathcal{RD}$ as illustrated in the last section and determines the $\mathcal{RAS}$ by mapping the objects in the *needlist* to $\mathcal{RAS}$ using $\mathcal{RAF}$ function. When the primary site pushes an update for a specific site and that data item is used by a local transaction, after committing the transaction, the site sends a need request (need req.) to the primary site which in turn add it to its need list. Note that, if the selected site use that data item one more time, it will not sends a need request unless the primary sends a new replica.

Algorithm 1 shows the pseudocode for the proposed replica control algorithm when a specific node receives a read transaction. And Algorithm 2 illustrates the steps when an update transaction is received at node N. When a read transaction is received at node N, the algorithm checks for an existence of active replica(s) by checking the validity interval of the required object(s), if it exists, it will be used by the transaction, otherwise a requested transaction is created and sends to the primary site containing that object(s). Note we assume that only one site can be requested by the transaction.

If an update transaction is received, a check for that if the requested object(s) is a primary object (Local, Shared) is maintained, as it was previously illustrated that the primary site is responsible for updating its local data objects. If the requested object(s) is a shared data object, a validity check is done, and the cooperation phase is done between this site and the object primary site. When the primary site receives the update request, it first checks the conflict existence using ($\mathcal{VN}$) of the object. And it then starts the propagation phase to other sites using the $\mathcal{RD}$ and $\mathcal{RAS}$ values generated by the Replication Manager. Transaction Manager must differentiate between the update transactions and the propagation transactions to avoid a cycle of endless propagation process, simply, a transaction type could be used.

## 4. Performance Evaluation

A full simulation environment have been developed to test the proposed allocation algorithm, we have chosen system parameter values that are typical of today's technology capabilities, e.g., network delays. The settings for the system parameters are given in table1, while the settings for user transaction are given in Table 2. A user transaction consists of operations on temporal data objects including both sensor and derived data objects.

The transactions arrival rate follows Poisson arrival pattern, the arrival rate ג varied from (10- 80) transactions per second, accordingly, the workload applied is approximately varied from (50%-100%) when the arrival rate varied from (10-80) respectively. The execution time for one operation is between 100 microseconds to 1000 microseconds, and the transaction execution time is exponentially distributed with mean (3). The sensor execution time is uniformly distributed between (0.1 – 1) second, and the slack factor of transactions is set to 5. The Remote Data Ratio is the ratio of the number of remote data operations (operations that access data hosted by other sites) to that of all data operations. The remote data ratio is set to 20%, which means 20 percent of transaction operations are remote data operations. At each node, the entire workload varied from 20-100%.
All simulation results are based on at least ten runs, to evaluate our algorithm we use no replication and full replication as two baseline protocols. These two algorithms are the simplest, but widely used replication control strategies. The transaction miss ratios and number of messages (reflects the network overhead) of the three algorithms are shown in Fig. 4. As we can see from the figure, among the three algorithms, the proposed algorithm gives the best transaction miss ratios under different transaction workloads.

Table 1: System Parameter Settings

| Parameter | Value |
|---|---|
| Node # | 10 – 50 |
| Network Delay | 1 – 3 ms |
| Temporal Derived Data # | 200/Node |
| Temporal Sensor Data # | 100/Node |
| Base Update Frequency | Uniform(0.1 - 1) sec |
| System Load | 20-100% |

Table 2: Transaction Parameter Settings

| Parameter | Value |
|---|---|
| Sensor Transaction # | 300 |
| User transaction # | 700 |
| Write Operation Time | 5 ms |
| Read Operation Time | 3 ms |
| Slack Factor | 5 |
| Remote Transaction ratio | 20% |
| Read/Write operation Prob. | (0.4 – 0.6 ) Respectively |
| Execution Time Of Sensor Tran. | Uniform (0.1 – 1) s |
| Execution Time Of user Tran. | Exep (3) |
| Execution Time Of Propagation Tran. | Exep (3) |
| Transaction Arrival Rate | (20 – 80) Trans/s |

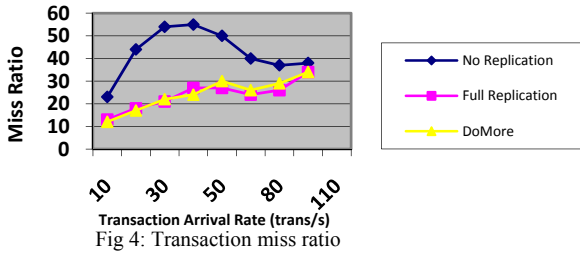


Fig 4: Transaction miss ratio

## 5. Conclusion and Future Work

In this paper, we present a dynamic replication control algorithm designed for medium and large scale distributed real-time database systems. The algorithm is designed to receive both periodic and aperiodic transactions and the system has no a prior knowledge of its data requirements. The replicas of the data items are being dynamically allocated to distributed nodes according to their access pattern. The model also allows different degree of consistency for each data object which is dynamically calculated according different. A detailed simulation study shows that our algorithm can greatly improves the system performance compared to the system without replication or system with simple full replication strategy. It is desirable to enhance and extend our algorithm to deal with transaction of many operations instead of one operation, and deal with other parameters that affect the performance issues for distributed real-time database, such as using one of the concurrency control protocol to enable concurrent execution of transactions.

## Appendix

```
//Define NeedList[n] array of nodes for each object o O : initially empty;
// LSiteId : the id of the local site where the transaction is initiated.
// RSiteId: the Id of the remote site to which the transaction is sent.
// O =(Id,Type,Name,PsiteId,Value,TS,VI,BUF,VN,FR)
// TRead :{ (Tid,LsiteId,Rsiteid,RS,D,RGUF,DL,e) has been submitted}
Begin
    If LSiteId = RsiteId //Check if the transaction is local transaction
        Then // check if the object is primary object
        If O(PsiteId) = RsiteId
        Then Return result from executing Tread on O;
        Commit (TRead);
        End
            Else If Current time – O(TS)<= O(VI)
            Then  //check if it is active replica;
            Return result from executing Tread on O;
            Commit (TRead);
            End
                Else Create A remote Read transaction
                Create local update transaction;
                Return result from executing Tread on O;
                Commit (TRead);
                End
        Else Return result from executing Tread on O;
        Commit (TRead);
      append( (LsiteId)) // Append ith Site  to the top of the needlist
End
```

Algorithm 1: Replica control algorithm on receiving a Read transaction

```
// Define NeedList[n]array of nodes for each object o O:initially empty;
// LSiteId:the id of the local site where the transaction is initiated.
// RSiteId:the Id of the remote site to which the transaction is sent.
// O =(Id,Type,Name,PsiteId,Value,TS,VI,BUF,VN,FR)
// Tupdate :{ (Tid ,LsiteId,Rsiteid,WS,RS,D,RGUF,DL,e)has been received}.
// Define RD int : replication degree  1< RD <N.
// Define  RAS[n] array of nodes for each object o O: initially empty;
// Define i int;

Begin
If LSiteId = RsiteId        //Check if the transaction is local
Then
If O(PsiteId)= RsiteId
Then // check if the object is primary object
      Return result from executing Tupdate on O;
      Commit (TRead);
      VN+1; // enter the propagation phase
      For j:=1  to RD do
      Perform RAF (Si (id))  // Append ith Site located in the needlist(o)  to the RAS
      End For
      For each site S in RAS Do
      Begin
      Create  Tupdate  (S(id));
      End
      End For;
        Else Return result from executing Tupdate on O;
        Receive Acknowledgment from the primary site.
        Commite Tupdate;
Else  //if the transaction is a remote update transaction ;
      Return result from executing Tupdate on O;
      If O(PsiteId) ≠ RsiteId Then // check if the opject is not a local object;
      Commite (Tupdate);
      End if
      VN(o)+1;
      Else
      Send Aknowleadgment to the Tupdate(Lsiteid);
      append(Si (id)) // Append ith Site  to the end of the needlist
            VN(o)+1;
            Commite (Tupdate); // enter the propagation phase
            For j:=1  to RD do
                  Perform RAF (Si (id))  // Append ith Site located in the needlist(o)  to
                  the RAS
            End for
            For each site S in RAS Do
                  Begin
                  Create  Tupdate  (S(id));
            End for
End
```

Algorithm 2: Replica control algorithm on receiving an update transaction

## References

[1] K. Ramamritham, 'Real-time databases', International Journal of Distributed and Parallel Databases 1(2),, pp. 199–226, 1993.

[2] J. Gray, P. Helland, P. O'Neil, D. Shasha, " The dangers of replication and a solution". In: Proc. of the ACM SIGMOD International Conf. On Management of Data, Vol. 25, No. 2 of ACM SIGMOD Record. ACM Press, 1996, pp. 173–182.

[3] F.B. Schneider, "Implementing Fault-Tolerant Services Using the State Machine Approach A Tutorial," ACM Computing Surveys, Vol.22, No.4, 1990, pp. 299-319.

[4] W. Matthias, S. André "Comparison of Database Replication Techniques Based on Total Order Broadcast " IEEE Trans. Knowledge Data Eng. Vol.17, No.4, 2005, pp. 551-566.

[5] F. Pedone, R. Guerraoui and A. Schiper "The Database State Machine Approach", Journal of Distributed and Parallel Databases and Technology, Vol.14, No.1,July 2003, pp. 71-98.

[6] J. Holliday, D. Agrawal, and A.E. Abbadi, "The Performance of Replicated Databases Using Atomic Broadcast Group Communication,"Technical Report TRCS99-11, Computer Science Dept.,Univ. of California, Santa Barbara, 1999.

[7] F. Pedone, "The Database State Machine and Group Communication Issues," PhD Thesis, EEcole Polytechnique Fe´de´rale de Lausanne, Switzerland, 1999..
[8] B. Kemme and G. Alonso, "A New Approach to Developing and Implementing Eager Database Replication Protocols," ACM Trans. Database Systems, vol. 25, no. 3, pp. 333-379, 2000.
[9] Y. Saito, M. Shapiro, Optimistic replication. ACM Comput. Surv. Vol.37, No.1, pp. 42-81, 2005
[10] S. Hyuk Son."Replicated data management in distributed database systems", SIGMOD Rec. Vol.17, No.4, 62-69, 1988
[11] ] O. Wolfson, S. Jajodia, Y. Huang, "An adaptive data replication algorithm", ACM on Database Systems (TODS) , Vol.22,No. 2, pp. 255-314, 1997
[12] ] B. Kemme, G. Alonso, " A Suite of Database Replication Protocols based on Group Communication Primitives", In Proceedings of the The 18th International Conference on Distributed Computing Systems (ICDCS '98). IEEE Computer Society, Washington, DC, USA, pp. 156-.1998
[13] M. Wiesmann, F. Pedone, A. Schipe, B. Kemme, G. Alonso," Understanding replication in databases and distributed systems", In: Proc. 20th International Conference on Distributed Computing Systems (ICDCS 2000), Taipei, Taiwan, R.O.C., pp. 264–274, 2000.
[14] S. Cook, J. Pachl, I. Pressman, "The optimal location of replicas in a network using a READ-ONE-WRITE-ALL policy", Distrib. Comput, Vol. 15, No.1, pp. 57-66, 2002.
[15] M. Xiong, K. Ramamritham, J. Haritsa, J. Stankovic, " MIRROR A state conscious concurrency control protocol for replicated real-time databases", In Proc. 5th IEEE Real-Time Technology and Applications Symposium (RTAS 99), pp. 100–110,1999.
[16] G. Mathiason, S. Andler," Virtual full replication: Achieving scalability in distributed real-time main-memory systems. In: Proc. of theWork-in-Progress Session of the 15th Euromicro Conf. on Real-Time Systems. (2003)
[17] J. Barreto, "Information sharing in mobile networks: a survey on replication strategies " Technical Report RT/015/03Instituto Superior T´ecnico/Distributed Systems Group, Inesc-ID Lisboa, 2003.
[18] Y. Wei, A. Aslinger, S. H. Son, J.A. Stankovic, " ORDER: A Dynamic Replication Algorithm for Periodic Transactions in Distributed Real-Time Databases", In Proceedings of Real-time and Embedded Computing Systems and Applications, pp.152—16, 2004.
[19] P. Peddi, L. DiPippo, "A replication strategy for distributed real-time object oriented databases", In: Fifth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing. (2002)
[20] Sanny Syberfeldt, "Optimistic Replication with Forward Conflict Resolution in Distributed Real-Time Databases", Dissertation No. 1150, Linköping 2007
[21] T. Haerder, A Reute, " Principles of transaction-oriented database recovery", ACM Comput. Surv, Vol.15, No.4, pp. 287—317, December 1983.
[22] M. Shapiro, K. Bhargavan, Y. Chong, Y. Hamadi "A formalism for consistency and partial replication", 2004.
[23] T. Gustafsson, "Maintaining Data Consistency in Embedded Databases for Vehicular Systems", Licentiate Thesis, Linköping Studies in Science and Technology Thesis No. 1138. Linköping University, 2004.
[24] S. Gustavsson, S.F. Andler , "Self-stabilization and eventual consistency in replicated real-time databases ", in Proceedings of the first workshop on Self-healing systems", (WOSS '02), Charleston, SC, USA, ACM, pp. 105—107, 2002.
[25] Gustavsson, S. F. Andler, S. (2005), Continuous consistency management in distributed real-time databases with multiple writers of replicated data", Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) - Workshop 2 – Vol. 03, 2005.
[26] S. Gustavsson, S. F. Andler, " Real-time conflict management in replicated databases", in 'Proceedings of the Fourth Conference for the Promotion of Research in IT at New Universities and University Colleges in Sweden (PROMOTE IT 2004), Karlstad, Sweden', Vol. 2, pp. 504–513, 2004.
[27] S. Andler, J. Hansson, J. Eriksson, J. Mellin, M. Berndtsson, B. Eftring, "DeeDS towards a distributed and active real-time database system", SIGMOD Record, Vol. 25, No.1, pp. 38–51, 1996.
[28] S. Andler, J. Hansson, J. Eriksson, J. Mellin, M. Berndtsson, B. Eftring, An overview of the DeeDS real-time database architecture" , in 'Proceedings of the Sixth International Workshop on Parallel and Distributed Real- Time Systems', 1998.
[29] J. Baulier, P., Bohannon, S., Gogate, C., Gupta, S., Haldar, "DataBlitz storage manager : Main memory database performance for critical applications", Proceedings SIGMOD '99 of the 1999 ACM SIGMOD international conference on Management of data, ACM SIGMOD Record, Vol.28, No. 2, pp. 19–520, 1999.
[30] G. Mathiason, "Segmentation in a distributed real-time main memory database", Master's thesis, University of Sk¨ovde, Sweden, 2002.
[31] R. Elmasri , S. Navathe, "Fundamentals of Database Systems", 6th Edition, Addison Wesley, 2010.
[32] C. Mohan, B.Lindsay, and R.Obermarck," Transaction management in the R* distributed database management system", ACM Trans. Database Syst. Vol.11, No.4 , pp.378-396, 1986.
[33] Torky Sultan, Hazem M. El-Bakry, Hala A. Hameed, " General Framework for Modeling Replicated Real-Time Database", International Journal of Electrical and Computer Engineering, Vol. 4, No. 8, pp. 505-511, 2009.
[34] D.L. McCue, M.C. Little, "Computing Replica Placement in Distributed Systems" A Position Paper for the Second Workshop on the Management of Replicated Data. University of Newcastle upon Tyne Appeared in the Proceedings of the IEEE Second Workshop on Replicated Data, Monterey, pp 58-61, 1992.

.

**T. Soltan** is Professor with Faculty of Computer Science and Information Systems – Helwan University, Helwan – Egypt.

**Hazem M. El-Bakry** (Mansoura, EGYPT 20-9-1970) received B.Sc. degree in Electronics Engineering, and M.Sc. in Electrical Communication Engineering from the Faculty of Engineering, Mansoura University – Egypt, in 1992 and 1995 respectively. Dr. El-Bakry received Ph. D degree from University of Aizu - Japan in 2007. Currently, he is assistant professor at the Faculty of Computer Science and Information Systems – Mansoura University – Egypt.His research interests include neural networks, pattern recognition, image processing, biometrics, cooperative intelligent systems and electronic circuits. In these areas, he has published more than 75 papers in major international journals and 150 papers in refereed international conferences.Dr. El-Bakry has the United States Patent No. 20060098887, 2006. Furthermore, he is associate editor for journal of computer science and network security (IJCSNS) and journal of convergence in information technology (JCIT). In addition, is a referee for IEEE Transactions on Signal Processing, Journal of Applied Soft Computing, the International Journal of Machine Graphics & Vision, the International Journal of Computer Science and Network Security, Enformatika Journals, WSEAS Journals and many different international conferences organized by IEEE. Moreover, he has been awarded the Japanese Computer & Communication prize in April 2006 and the best paper prize in two conferences cited by ACM. He has also been awarded Mansoura university prize for scientific publication in 2010 and 2011. Dr. El-Bakry has been selected in who Asia 2006 and BIC 100 educators in Africa 2008.

**H. Abdel hameed** is assistant lecturer with Faculty of Information Technology – Misr University for Science and Technology – Al-Motamayez District 6th of October City – Egypt

# Performance Evaluation of QoS Parameters in Dynamic Spectrum Sharing for Heterogeneous Wireless Communication Networks

**Kaniezhil. R[1] and Chandrasekar. C[2] and NithyaRekha.S[3]**

**[1,3] Research Scholar, Department of Computer Science, Periyar University, Salem, TamilNadu-636011, India**

**[2] Associate Professor, Department of Computer Science, Periyar University, Salem, TamilNadu-636011, India**

**Abstract**

Cognitive radio nodes have been proposed as means to improve the spectrum utilization. It reuses the spectrum of a primary service provider under the condition that the primary service provider services are not harmfully interrupted. A cognitive radio can sense its operating environment's conditions and it is able to reconfigure itself and to communicate with other counterparts based on the status of the environment and also the requirements of the user to meet the optimal communication conditions and to keep quality of service (QoS) as high as possible. The efficiency of spectrum sharing can be improved by minimizing the interference. The Utility function that captures the cooperative behavior to minimize the interference and the satisfaction to improve the throughput is investigated. The dynamic spectrum sharing algorithm can maintain the quality of service (QoS) of each network while the effective spectrum utilisation is improved under a fluctuation traffic environment when the available spectrum is limited.

***Keywords:*** *CR, throughput, propagation delay, spectrum efficiency, Interference.*

## 1. Introduction

It is commonly believed that there is a spectrum scarcity at frequencies that can be economically used for wireless communications. By recent studies of FCC, it shows that the scared spectrum can be well utilized and unused spectrum ie 'white spaces' can be utilized by the secondary users with the advance technology of Cognitive Radio (CR)[2] to implement the opportunistic spectrum sharing.

However, as noted by the FCC, there are large portions of allotted spectrum that are unused when considered on a time and geographical basis. There are portions of assigned spectrum that are used only in certain geographical areas and there are some portions of assigned spectrum that are used only for brief periods of time. Studies have shown that even a straightforward reuse of such "wasted" spectrum can provide an order of magnitude improvement in available capacity.

Thus the issue is not that spectrum is scarce – the issue is that most current radio systems do not utilize technology to effectively manage access to it in a manner that would satisfy the concerns of current licensed spectrum users. Cognitive radio [2], [3], [4] is currently considered as one of the most promising solutions to the aforementioned scarcity problem by enabling a highly dynamic, device-centric spectrum access in future wireless communication systems.

A CR can adapt the operation parameters of its radio (frequency band, modulation, coding etc) and its transmission or reception parameters on the fly based on cognitive interaction with the wireless environment in which it operates. CR will lead to a revolution in wireless communication with significant impacts on technology as well as regulation of spectrum usage to overcome existing barriers.

Cognitive radio not only adapts to the available spectrum but it also shows the better Qos and the channel conditions that satisfies the requirement of the effective performance of the bandwidth.

Cognitive Radio is an emerging technology provides an way to efficient way for better utilization of the unused spectrum. Spectrum allocated to the primary users is not used fully at all instances of times. Hence, the number of trying to use this unused licensed spectrum is increasing enormously. So, the idea is that the sensing the unused or empty frequencies of the primary users and that can be

accommodated to some other unlicensed (Secondary) users. This makes the efficient utilization of the available spectrum. This can be achieved by using the Cognitive radio to identify and used to allocate the unused spectrum bandwidth that can allocate dynamically by changing their parameters keeping in view the QoS requested by the secondary user or simply the application, without interfering with the primary users.

The techniques developed to date for the enhancement of heterogeneous networks concentrate on improving their accessibility and QoS. Numerical simulation results demonstrate that Throughput and Spectrum Efficiency of networks employing dynamic spectrum sharing are much better than those of networks employing fixed allocation, especially for networks under heavy traffic load, when spectrum is limited.

I have already proposed the spectral efficiency in my previous work that the call arrival rate vs spectral efficiency[1]. The remaining work ie Performance of QoS based on spectrum sharing using CR nodes is carried out in the present work.

The paper is organized as follows; Section 2 and 3 defines cognitive radio and proposes a system model approach for its implementation. In section 4, Performance analysis has been investigated to improve the system efficiency. Section 5 and 6 presents the Proposed Algorithm of the work and simulation results with implementation issues. Finally, conclusions are presented in Section 7.

## 2. Cognitive Radio

### 2.1 Introduction

A "Cognitive Radio" is a radio that is able to sense the spectral environment over a wide frequency band and exploit this information to opportunistically provide wireless links that best meet the user communications requirements. CR provides the real time interaction with its environment. This provides the way to dynamically adapt to the dynamic radio environment and the radio analyzes the spectrum characteristics and changes the parameters among the users that share the available spectrum. With the approach to solve the issue of scarcity of available radio spectrum, the Cognitive radio technology is getting a significant attention [4]-[6].

The primary feature of cognitive radio is the capability to optimize the relevant communication parameters given a dynamic wireless channel environment. Since cognitive radios are considered lower priority or secondary users of spectrum allocated to a primary user, a fundamental requirement is to avoid interference to potential primary users in their vicinity. On the other hand, primary user networks have no requirement to change their infrastructure for spectrum sharing with cognitive networks.

Therefore, cognitive radios should be able to independently detect primary user presence through continuous spectrum sensing. In general, cognitive radio sensitivity should outperform primary user receiver by a large margin in order to prevent what is essentially a *hidden terminal problem*. This is the key issue that makes spectrum sensing very challenging research problem.

### 2.2 Cognitive Radio Parameters

The Cognitive Radio system must relate the performance objectives to the transmission parameters and the environmental parameters in order to reach at an optimized solution. While defining the list of parameters we make a compromise between the large time scale, system level parameters and the small time scale, transmission level parameters.

Table I shows the transmission parameters used in this paper to generate a utility function.

TABLE I

TRANSMISSION PARAMETER LIST

| ***Parameter Name*** | ***Symbol*** | ***Description*** |
|---|---|---|
| Transmit Power | P | Transmission Power |
| Modulation Type | MT | Type of Modulation |

The available system parameters should be defined as decision variables for evolutionary algorithms calculating generating utility functions. Table II shows the Environmental parameters used in this paper to generate a utility function.

The BER parameter value depends on several channel characteristics, including the noise level and transmit power. Environmental Parameters inform the system of the surrounding environmental characteristics. SBAC Algorithms is chosen for the allocation algorithm due to their fast convergence.

TABLE II

ENVIRONMENTALLY SENSED PARAMETER LIST

| *Parameter Name* | *Symbol* | *Description* |
|---|---|---|
| Bit Error Rate | BER | Number of bit errors divided by the total number of transferred bits during a studied time interval. |
| Signal-to-Interference Noise Ratio | SINR | Ratio of the received strength of the desired signal to the received strength of undesired signals (noise and interference). |
| Noise power | $N_O$ | Magnitude in decibels of the Noise Power |

### 2.3 Utility Functions

The system performance indexes are described in terms of utility functions. The actual results should take balance of these utility functions, which can meet the QoS requirements and improve the performance.

Utility functions are defined individually considering the current user's QoS specifications. This implies to the existence of a trade-off among the parameters for a particular channel. This is analyzed by the corresponding weights assigned by the user to each of them. This is actually very useful in our decision-making process and provides with a variety of solutions for the best optimization of a problem.

Four performance measures of Data Transmission rate, Propagation Delay, Spectral Efficiency and Throughput are considered in this paper and the utility functions are designed as in Table III:

TABLE III

UTILITY FUNCTIONS

| *Performance Metrics* |
|---|
| Low Propagation Delay |
| Minimize RTT |
| Maximize Throughput |
| Maximize Spectral Efficiency |

Using the objectives in Table III as sole inputs to the utility functions will not suffice. It is ambiguous to have the system minimize power consumption while also minimizing BER. Thus, the objectives must also contain a quantifiable rank representing the importance of each. This will allow the utility function to characterize the trade-offs between each objective by ranking the objectives in order of importance. Several approaches exists for determining the preference information of a set of objectives.

## 3. System Model

We consider the spectrum sharing among multiple service providers, they belong to the licensed bands. We assume that there are a number of primary and secondary users communicating with their partners simultaneously. Here, the term "user" will be used broadly where it can be a mobile node or base station in a distributed networks. Simultaneous communications among users (i.e., both primary and secondary users) will interfere with each other.

The entities we will work with are communication links each of which is a pair of users communicating with each other. We will refer to communication links belonging to secondary networks as secondary links. We will also consider the interference constraints at the receiving nodes of primary networks which will be referred to as primary receiving points. We assume that each primary receiving point can tolerate a maximum interference level. Also, secondary links have desired QoS performance in terms of BER.

We assumed a model in which $S$ base-stations are sharing $S$ different frequency bands. Each band has a user capacity of $K$ with throughput of $R$ per user. Therefore, each band can support an aggregate traffic of $KR$ bits per second per Hertz. In theory, each base-station achieves this throughput via the licensed band provided that there are $K$ active users and enough packets from each user to fully exploit the capacity.

We should note at this point that the proposed sharing protocol applies to both downlink and uplink transmissions. The access strategy used by each base-station to serve the users could be any of the standard techniques.

Active users are defined to be users of the wireless network requesting access for their data flow. Assume each base station $i \varepsilon \{1, 2, \ldots, S\}$ has a random number of requests for establishing a session say $ai$ each distributed according to a Poisson distribution with average rate of $\lambda$. Each $ai$ is assumed to be independent from requests at other stations. We assume that all users and sessions have the same data rate requirements met by the rate $R$.

## 4. Performance Analysis

Assume that there are $M$ primary receiving points and $N$ secondary communication links in the considered geographical area. Let us denote the channel gain from the

transmitting node of secondary link *i* to receiving node of secondary link *j* by $g^{s}_{(j,i)}$ while the channel gain from the transmitting node of secondary link *i* to primary receiving point *j* as $g^{p}_{(j,i)}$ .

If $N_i$ denotes the total noise and interference at the receiving side of secondary link *i*, for wireless access system, the corresponding effective bit-energy-to-noise spectral density ratio can be written as

$$\mu_i = \frac{W}{R_i} \frac{g^{s}_{(j,i)} P_i}{\sum_{j=1, j\neq i}^{N} g^{s}_{(j,i)} P_j + N_i}$$

Where $W$ is the spectrum bandwidth, $R_i$ is the transmission rate of secondary link *i*. Here, $W / R_i$ is the processing gain which is usually required to be larger than a particular value. The processing gain is simply equal to one for other multiple access technologies and $\mu_i$ denotes the SINR. Now, if a particular modulation scheme is employed, there will be an explicit relation between BER and SINR.

Thus, for a specific required BER level of secondary link *i*, $\mu_i$ is required to be larger than a corresponding value $\gamma_i$ . Hence, the QoS requirement for secondary link *i* can be expressed as

$$\mu_i \geq \gamma_i, i=1,2,...,N$$

Now, let $T_j$ be the minimum interference level tolerable by primary receiving point j. The interference constraint for primary receiving point *j* can be written as

$$\sum_{i=1}^{N} g^{p}_{(j,i)} P_i \leq T_j, j=1,2,...,M$$

where total interference at the primary receiving point *j* should be smaller the tolerable limit.
We will assume that transmission rate of secondary link *i* can be adjusted in an allowable range with minimum and maximum values are $R_i^{\min}$ , $R_i^{\max}$ respectively. Also, power of secondary link *i* is constrained to be smaller than the maximum limit $P_i^{\max}$ .

## 5. Proposed Algorithm

The proposed work, QoS of Spectrum sharing among multiple Service Providers is carried out in a long-term spectrum Assignment scheme. The function coordinates and negotiates the spectrum assignments between multiple Service Providers for large geographical areas.

Algorithm : SBAC

```
BS ← mn   (mobile node send request to Base Station)
CR ← BS
nCR ← CR
current _channel _available _list ← nCR
prob ← current _channel _available / total _channel
∀ ch_frq
    if frmax<ch_frq
        frmax ← frq
    end
    if frmin>ch_frq
        frmin ← ch_frq
    end
end
inter ←| frmax − frmin |
cost ← t * 60 * c
ch _u = (10 * β1 * prob) + β2 * log 1/inter + β3 * 1/cost
cu _list[cu _count++] =| ch _u |
if (cu _count != 0)
    ∀cu _list
        if maxi < cu _list
            maxi = ch _list
        end
    end
channel _maxi ← maxi
end
```

The spectrum assignments are updated periodically and it is explained with the help of the proposed algorithm named SBAC (Selection of Best Available Channel).

## 6. Simulation Results

The Cognitive Radio receives the RF environment at its receiver and involves itself in a decision-making process to accommodate a new user requesting the spectrum allocation. This requires a decision-making considering certain factors, such as the secondary user's requirements as parameters like, its Channel coding, data transmission rate, etc. The user needs the spectrum to carry out its communications and specifies its QoS requirements to the cognitive radio that also gets the information about the RF environment from a sensing module.

The utility function represents the radio's behavioral traits for the decision-making process to achieve the required optimization. There can be many possible traits that can be considered in this regard but we shall consider only some of the basic traits for the radio in this research. Some of the possible traits that can be considered are the occupied bandwidth, spectral efficiency, throughput and delay.

We shall just consider a few parameters only, in order to maintain the simplicity in the research. These are the frequency bands, power and BER.

### 6.1 Minimize Propagation Delay

Propagation delay is the amount of time taken for the signal to travel from the sender to the receiver over a medium. It can be computed as the ratio between the link length and the propagation speed over the specific medium.

$$\text{Propagation delay} = d / s$$

where d is the distance travelled and s is the propagation speed.

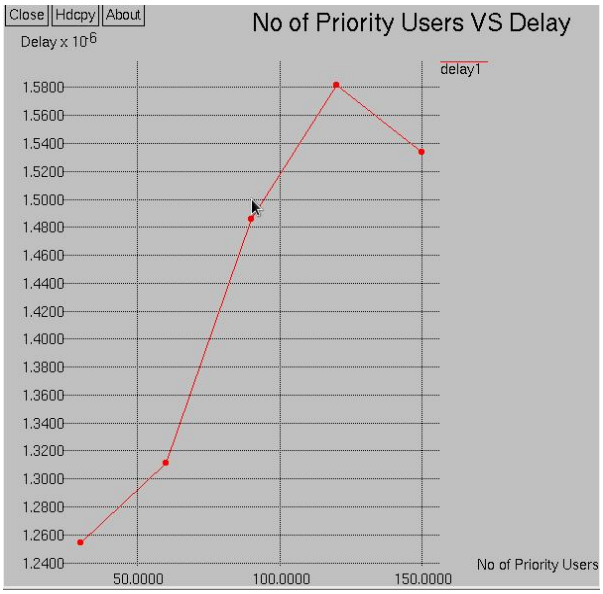



Fig : 1 No. of users VS Propagation Delay

In the proposed work, as the number of Users increases propagation delay gets minimized as shown in the Fig : 1.

### 6.2 Maximize Throughput

The maximum data rates of the TX and RX depend on the bandwidth of the circuits, and simplistically it might seem that the least of these will settle the issue.

But for communication between them the max data rate is affected by the noise in the system, and this will depend on the noise of the propagation medium, the noise figure of the RX, the power level of the TX, the transmission loss, and the maximum tolerable error rate.

Satisfactory data transmission can be achieved with higher noise at a lower bit rate because of the statistical nature of the noise, and the time-domain averaging of signals which occurs in the RX.

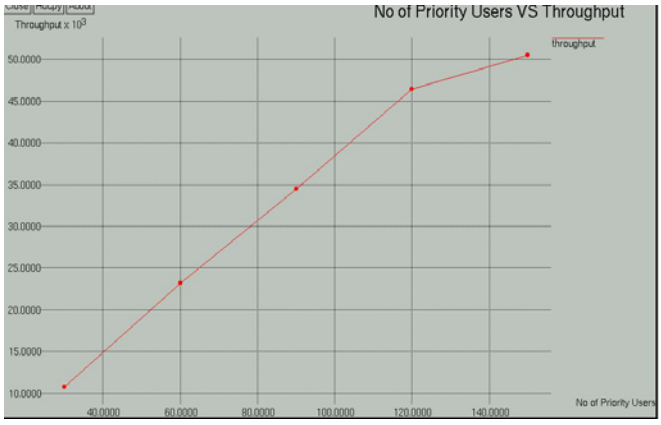



Fig : 2 No. of Users VS Throughput

The objective of the paper represents that improving the communications quality of the radio. Maximizing the throughput deals with the data throughput rate of the system. Emphasizing this objective, the overall system throughput should be increased and it is reached in the proposed work as shown in the Fig : 2. This refers to the increase in overall data throughput transmitted by the signal.

### 6.3 Minimize RTT

Round-trip time (RTT) is the time it takes for a client to send a request and the server to send a response over the network, not including the time required for data transfer.

Current Round-Trip Time (RTT) of every active connection is estimated in order to find a suitable value for the retransmission time-out.
RTT is the major contributing factor to latency on "fast" (broadband) connections and it's especially important to minimize the number of requests that the client needs to make and to parallelize them as much as possible.

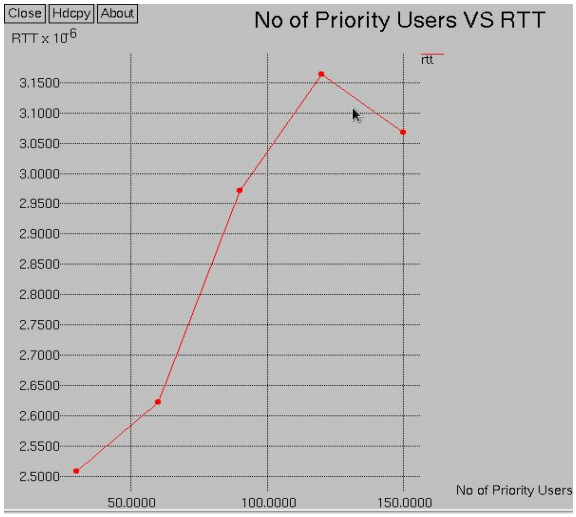



Fig : 3 No. of Users VS RTT

In the proposed work, as the number of priority users increases the RTT value decreases, as to minimize the number of round trips that need to be made.

### 6.4 Minimize Interference

Interference is the key factor that limits the performance of wireless networks. The problem may be thought of as arising from the limitations of the receiver: better receivers are more able to extract the desired signal from a noisy environment of background radiation and other transmitters.

$$interference = \mid frmax - frmin \mid$$

In this paper, an interference aware dynamic spectrum sharing method is applied. It dynamically minimizes the inter-cell interference and significantly improves the system performance.

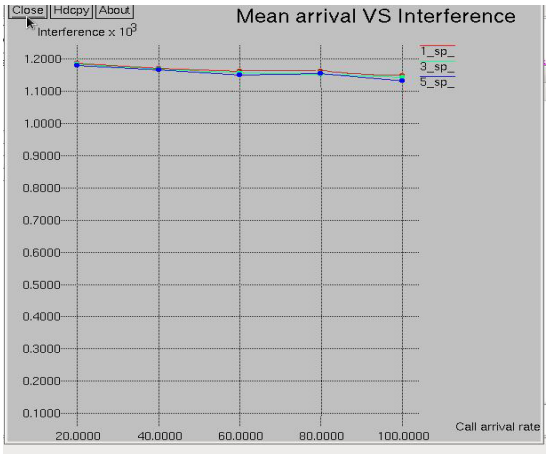



Fig : 4 Mean Call Arrival VS Interference

As the mean call arrival increases the Interference gets decreased and there will be a minute variations in the Interference as shown in the Fig : 4.

### 6.5 Maximize Spectral Efficiency(η)

The Spectrum Efficiency $\eta_s$ is the ratio of average busy channels over total channels owned by service providers ie it refers to the amount of information that can be transmitted over a given bandwidth.

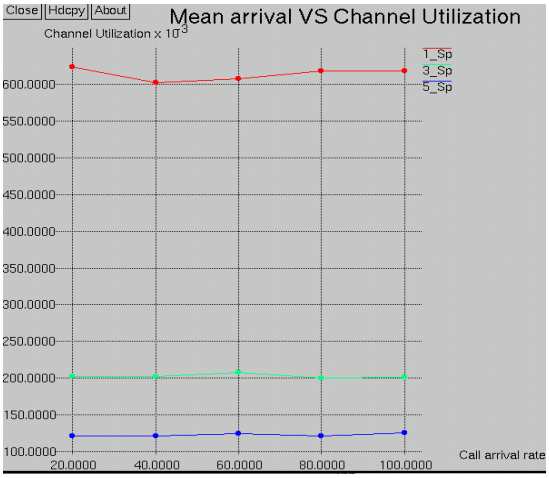



Fig : 5 Mean Call Arrival VS Spectrum Efficiency

As the mean call arrival increases the channel Utilization also increases as shown in the Fig :5. Higher Spectrum efficiency is estimated because the call blocking rate is lower; thus more calls can contribute to the spectrum utilization

## 7. Conclusions

The proposed dynamic spectrum sharing algorithm has been shown to be an effective solution for improving the spectrum efficiency under fluctuating traffic loads while maintaining the Interference and throughput in their acceptable QoS levels. It is illustrated that this model can be successfully employed in the key spectrum allocation decisions in such a spectrum sharing environment in a heterogeneous wireless network.

In this paper, we have studied the distribution of the interference generated by a secondary network to a primary network. We have derived a general formula for the interference taking into account the cognitive ability, throughput and transmit power. Also, Cognitive radio parameters, Utility Functions, QoS and interference constraint parameters on network performance are investigated and discussed.

## References

[1] R.Kaniezhil, Dr.C.Chandrasekar, S.Nithya Rekha, "Channel Selection for Spectrum Sharing using CR Nodes", International Proceedings of Computer Science and Information Technology, IACSIT Press, 2011, vol.20, pp 93-98.

[2] Danijela Cabric, Shridhar Mubaraq Mishra, Robert W. Brodersen, "Implementation Issues in Spectrum Sensing for Cognitive Radios" in Proc. 38th Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA, Nov. 2004, pp.772–776.

[3] F. Akyildiz, W. Y. Lee, M. C. Vuran, S. Mohanty, "NeXt generation dynamic spectrum access cognitive radio wireless networks: A survey," Computer *Networks* Journal (Elsevier), 2006, Vol. 50, pp. 2127-2159.

[4] Lars Berlemann, George Dimitrakopoulos, Klaus Moessner, Jim Hoffmeyer, "Cognitive Radio and Management of Spectrum and Radio Resources in Reconfigurable Network", Wireless World Research Forum, 2005.

[5] Jon M. Peha, "Emerging Technology and Spectrum Policy Reform", International Telecommunications Union (ITU) Workshop on Market Mechanisms for Spectrum Management, ITU Headquarters, Geneva, January 2007.

[6] R. Etkin, A. Parekh, and D.Tse, "Spectrum sharing for Unlicensed bands," in IEEE International Symposium on New Frontiers in Dynamic Spectrum Access, 2005, pp 251-258.

[7] T.J. Harrold, L.F. Wang, M.A. Beach, G. Salami, A. Yarmohammad, O. Holland, "Spectrum Sharing and Cognitive Radio Opportunities for Efficiency Enhancement", IEEE, 2009.

[8] Ammar Alshamrani, Xuemin (Sheman) Shen, and Liang Xie, "QoS Provisioning for Heterogeneous Services in Cooperative Cognitive Radio Networks", IEEE Journal on selected areas in Communications, April 2011,Vol. 29, no. 4.

[9] Leila Musavian and Sonia Assa, "Quality-of-Service Based Power Allocation in Spectrum-Sharing Channels" in the IEEE "GLOBECOM" 2008 proceedings.

[10] Yiping Xing, Chetan N. Mathur, M. A. Haleem, R. Chandramouli and K.P. Subbalakshmi, "Real-Time Secondary Spectrum Sharing with QoS Provisioning", IEEE CCNC, 2006.

[11] Alireza Attar, Oliver Holland, Mohammad Reza Nakhai, "Interference Management in Shared Spectrum for WiMAX Systems" in IEEE Proc. VTC Spring, pp.1620-1624, 2008.

[12] Kevin Fall, Kannan Varadhan., "The ns Manual", The VINT Project, May 9, 2010.

[13] Teerawat Issariyakul, Ekram Hossain, "Introduction to Network Simulator NS2" ISBN: 978-0-387-71759-3 e-ISBN: 978-0-387-71760-9, Springer 2009.

**R. Kaniezhil** is a member of IEEE. She is a Research scholar in the Department of Computer Science, Periyar University, Salem. She received her B.Sc Degree from University of Madras in 1998. She received her MCA and M.Phil Degrees from Periyar University and Annamalai university, in 2001 and 2007, respectively. Her research interests include Mobile computing, Spectrum and Wireless Networking.

**Dr. C. Chandrasekar** is a member of IEEE. He received his Ph.D degree from Periyar university. He is working as an Associate Professor, Department of Computer Science, Periyar University, Salem. His areas of interest include Wireless networking, Mobile Computing, Computer Communications and Networks. He is a research guide at various universities in India. He has published more than 40 technical papers at various National & International conferences and 43 journals.

**S.Nithya Rekha** is a member of IEEE. She is a Research scholar in the Department of Computer Science, Periyar University, Salem. She received her B.Sc Degree from Bharathiayar University in 1994. She received her MCA and M.Phil Degrees from IGNOU and PRIST university, in 2006 and 2008, respectively. Her research interests include Mobile Computing, Rough set and Wireless networking.

# Self Organizing Map -based Document Clustering Using WordNet Ontologies

**Tarek F. Gharib[1,2], Mohammed M. Fouad[3], Abdulfattah Mashat[1], Ibrahim Bidawi[1]**

**[1] Faculty of Computing and Information Technology, King Abdulaziz University Jeddah, Saudi Arabia**
**[2]Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt**

**[3] Faculty of Informatics and Computer Science, The British University in Egypt (BUE) Cairo, Egypt**

**Abstract**

With the rapid development of web content, retrieving relevant information is difficult task. The efficient clustering algorithms are needed to improve the results of the retrieval. Document clustering is a process of recognizing the similarity or dissimilarity among the given objects and forms subgroups sharing common characteristics**.** In this paper, we propose a semantic text document clustering approach that using WordNet lexical and Self Organizing Maps. The proposed approach uses the WordNet to identify the importance of the concepts in the document. The SOM is used to cluster the document. We use this approach to enhance the effectiveness of document clustering algorithms. The approach takes the advantages of the semantics available in knowledge base and the relationship between the words in the input documents. Some experiments are performed to compare efficiency of the proposed approach with the recently reported approaches. Experiments show advantage of the proposed approach over the others.

***Keywords:*** *Text Document Clustering; WordNet Lexical Categories; Self Organizing Map (SOM)*

## 1. Introduction

With the recent growth and diversity of electronic data on the World Wide Web (www), it becomes more difficult for Internet users to find the useful information from these huge amounts of data. Search engines and recommender systems help people to reduce the information overload by finding relevant information on their search topic. Clustering of documents is one of the techniques used in search engines and in recommender systems for efficiently finding documents that have similar topics [1], for improving the performance of information retrieval systems [2], for assisting users on a web site [3] and for personalization of search engine results [4]. Formally, document clustering is an optimization problem where the input of the problem is a set of documents and a (dis)similarity measure between these documents. Thus, similarity plays an important role in document clustering.

Text document clustering provides an effective navigation mechanism to organize this large amount of data by grouping their documents into a small number of meaningful classes. Text document clustering can be defined as the process of grouping of text documents into semantically related groups[5]. Most of the current methods for text clustering are based on the similarity between the text sources. The similarity measures work on the syntactically relationships between these sources and neglect the semantic information in them. By using the vector-space model in which each document is represented as a vector or 'bag of words', i.e., by the words (terms) it contains and their weights regardless of their order [6].

Vector space model is a popular model for document representation in document clustering including the above methods. Documents are represented by vectors of weights, where each weight in a vector denotes importance of a term in the document. In the standard VSM, however, semantic relations between terms are not taken into account. Two terms with a close semantic relation and two other terms with no semantic relation are both treated in the same way. This unconcern about semantics could reduce quality of the clustering result.

Many well-known methods of text clustering have two problems: first, they don't consider semantically related words/terms (e.g., synonyms or hyper/hyponyms) in the document. For instance, they treat {Vehicle, Car, and Automobile} as different terms even though all these words have very similar meaning. This problem may lead to a very low relevance score for relevant documents because the documents do not always contain the same forms of words/terms.

Second, on vector representations of documents based on the bag-of-words model, text clustering methods tend to use all the words/terms in the documents after removing the stop-words. This leads to thousands of dimensions in

the vector representation of documents; this is called the "Curse of Dimensionality". However, it is well known that only a very small number of words/terms in documents have distinguishable power on clustering documents and become the key elements of text summaries. Those words/terms are normally the concepts in the domain related to the documents [7].

There are some approaches that employ WordNet based semantic similarity to enhance the performance of document clustering [8, 9]. They modified the VSM model by readjusting term weights in the document vectors based on its relationships with other terms co-occurring in the document.

In this paper, we propose a semantic text document clustering approach that using WordNet lexical and Self Organizing Maps. The proposed approach uses the WordNet to identify the importance of the concepts in the document. The SOM is used to cluster the document. We use this approach to enhance the effectiveness of document clustering algorithms. The clustering performances are evaluated versus K-means and bisecting k-means algorithms. The approach takes the advantages of the semantics available in knowledge base and the relationship between the words in the input documents. Some experiments are performed to compare efficiency of the proposed approach with the recently reported approaches. Experiments show advantage of the proposed approach over the others.
The rest of this paper is organized as following; recent related work is discussed and presented in section 2. In section 3, we show the proposed semantic text clustering approach. In section 4 a set of experiments is presented to compare the performance of the proposed approach with current text clustering methods. Finally, conclusion and future work are given in section 5.

## 2. Related Work

In the recent years, text document clustering has been introduced as an efficient method for navigating and browsing large document collections and organizing the results returned by search engines in response to user queries [10]. Many clustering techniques are proposed like bisecting k-means [11], FTC and HFTC [12] and many others. From the performed experiments in [11] bisecting k-means overcomes all these algorithms in the performance although FTC and HTFC allows to reduce the dimensionality if the data when working with large datasets.

WordNet is used by Green [13-14] to construct lexical chains from the occurrences of terms in a document: WordNet senses that are related receive high higher weights than senses that appear in isolation from others in the same document. The senses with the best weights are selected and the corresponding weighted term frequencies constitute a base vector representation of a document.

Dave and Lawrence [15] use WordNet synsets as features for document representation and subsequent clustering. But the word sense disambiguation has not been performed showing that WordNet synsets decreases clustering performance in all the experiments. Hotho et al. use WordNet in an unsupervised scenario taking into account the WordNet ontology and lexicon. They used some strategy for word sense disambiguation which achieved improvements for the clustering results [16].

In [9] the authors explore the benefits of partial disambiguation of words by their PoS and the inclusion of WordNet concepts; they show how taking into account synonyms and hypernyms, disambiguated only by PoS tags, is not successful in improving clustering effectiveness because the noise produced by all the incorrect senses extracted from WordNet. Adding all synonyms and all hypernyms into the document vectors seems to increase the noise.

Reforgiato[17] presented a new unsupervised method for document clustering by using WordNet lexical and conceptual relations .In this work, Reforgiato uses WordNet lexical categories and WordNet ontology in order to create a well structured document vector space whose low dimensionality allows common clustering algorithms to perform well. For the clustering step he has chosen the bisecting k-means and the Multipole tree algorithms for their accuracy and speed.
Friedman et al. [18] introduced FDCM algorithm for clustering documents that are represented by vectors of variable size. The algorithm utilizes fuzzy logic to construct the cluster center and introduces a fuzzy based similarity measure which provided reasonably good results in the area of web document monitoring.

Hung and Wermter [19] proposed three novel text vector representation approaches for neural network based document clustering. The first is the extended significance vector model (ESVM), the second is thehypernym significance vector model (HSVM) and the last is the hybrid vector space model (HyM). ESVM extracts the relationship between words and their preferred classified labels. HSVM exploits a semantic relationship from the WordNet ontology. HyM is a combination of a TFxIDF vector and a hypernym significance vector, which combines the advantages and reduces the disadvantages from both unsupervised and supervised vector representation approaches. According to their experiments,

the self-organizing map (SOM) model based on the HyM text vector representation approach is able to improve classification accuracy and to reduce the average quantization error.

Sridevi and Nagaveni [20] proposed a model by combining ontology and optimization technique to improve the clustering. The proposed model uses the ontology similarity in identifying the importance of the concepts in the document. The particle swarm optimization is used to the cluster the document.

## 3. Semantic Text Document Clustering

In this section we describe in details the components of the proposed semantic text clustering approach. There are two main processes: Document Preprocessing that generated output document vectors from input text documents using WordNet[1] lexical information is introduced in the first step. The second step is Document Clustering that applies SOM neural network on the generated document vectors to obtain output clusters as illustrated in fig. 1.

### 3.1 Document Preprocessing

The first step in the proposed approach is document preprocessing which aims to represent the corpus (input documents collection) into vector space model. Data preprocessing is a very important and essential phase in an effective document clustering. The first part of feature extraction is preprocessing the lexicon and involves removal of stop words and stemming [6]. The stop words removal accounts to 20% to 30% of total words counts while the process of stemming reduces the number of terms in the document. Both the process helps in improving the effectiveness and efficiency of text processing as they reduce the indexing file size.

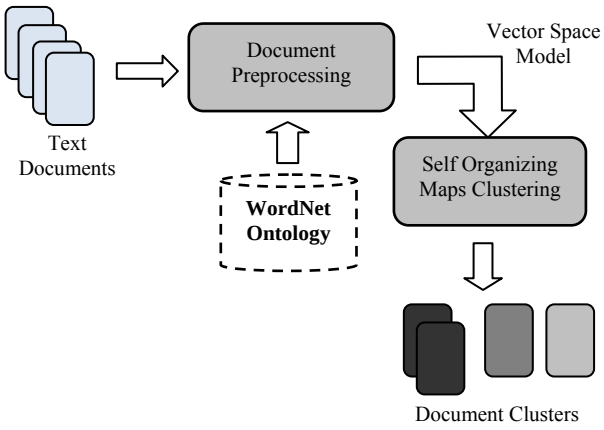



Fig. 1 Diagrammatic representation of the proposed approach

#### 3.1.1 Stopwords Removal

This is the first step in preprocessing which will generate a list of terms that describes the document satisfactorily. The document is parsed through to find out the list of all the words. The next process in this step is to reduce the size of the list created by the parsing process, generally using methods of stop words removal.

#### 3.1.2 Stemming

Stemming is process of linguistic normalization in which the variant forms of a word is reduced to a common form. For example: the word, connect has various forms such as connect, connection, connective, connected, etc., Stemming process reduces all these forms of words to a normalized word connect. Porter's English stemmer algorithm is used to stem the words for each of the document in our stemming process. This step aims to reduce the extracted frequent word list to optimize the next step for WordNet mapping. In our implementation we use minimum support value set to 10%, which means that the words found in less than 10% of the input documents is removed from the extracted word list.

#### 3.1.3 WordNet Lexical Category Mapping

As proposed in [17], we use WordNet lexical categories to map all the stemmed words in all documents into their lexical categories. We use WordNet 2.1 that has 41 lexical categories for nouns and verbs as shown in tables 1 and 2. For example, the word "dog" and "cat" both belong to the same category "noun.animal". Some words also has multiple categories like word "Washington" has 3 lexical categories (noun.location, noun.group, noun.person) because it can be the name of the American president, the city place, or a group in the concept of capital.

[1] WordNet project: http://wordnet.princeton.edu/

Some word disambiguation techniques are used to remove the resulting noise added by multiple categories mapping which are: disambiguation by context and concept map which are discussed in details in [13].

Table 1: WordNet nouns lexical categories

| | | | |
|---|---|---|---|
| Act | Animal | Artifact | Attribute |
| Body | Cognition | Communication | Event |
| Feeling | Food | Group | Location |
| Motive | Object | Person | Phenomenon |
| Plant | Possession | Process | Quantity |
| Relation | Shape | State | Substance |
| Time | Tops | | |

Table 2: WordNet verbs lexical categories

| | | | |
|---|---|---|---|
| Body | Change | Cognition | Communication |
| Competition | Creation | Contact | Perception |
| Emotion | Motion | Weather | Consumption |
| Social | Stative | Possession | |

The output vectors that are generated based on the number of words that found in each lexical category. The generated document vector D for each document d in the input text document is defined as in Eq. (1).

$$D^T = [X_1, X_2, ....., X_{41}] \quad (1)$$

We calculate $X_i$ as the number of words in document d that belongs to the $i^{th}$ lexical category in the WordNet lexical categories for the output vector.

## 3.2 Document Clustering

Clustering is one technology to find intrinsic structures in data sets. Text clustering method usually uses the document vector space model to split the document into vectors in high dimensional space, and then make clustering of these vectors. Text clustering can generally be divided into partitioned clustering algorithms and hierarchical clustering algorithms.

After generating the documents' vectors for all the input documents using feature extraction process, we start the clustering process as shown in fig. 1.

The problem of document clustering is defined as follows. Given a set of n documents called DS, DS is clustered into a user-defined number of k document clusters D1, D2,…Dk, (i.e. {D1, D2,…Dk} = DS) so that the documents in a document cluster are similar to one another while documents from different clusters are dissimilar.
In this stage we apply three different clustering algorithms which are k-means (partitioning clustering), bisecting k-means (hierarchical clustering) and SOM neural network. These algorithms are most commonly used in the document clustering step in the recent researches.

### 3.2.1 K-means and Bisecting k-means

We have implemented the k-means and bisecting k-means algorithms as introduced in [11]. We will state some details on bisecting k-means algorithm that begins with all data as one cluster then perform the following steps:

Step1: Choose the largest cluster to split.
Step2: Use k-means to split this cluster into two sub-clusters. (Bisecting step)
Step3: Repeat step2 for some iterations (in our case 10 times) and choose the split with the highest clustering overall similarity.
Step4: Go to step1 again until the desired k clusters are obtained.

### 3.2.2 Self Organizing Maps (SOM)

Self-organizing maps (SOM) learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on.

In this paper we focus on using SOM to perform the document clustering. The two reasons for using SOM rather than other clustering methods are that it is topologically preserving and clustering is performed non-linearly on the given input data sets. The topologically preserving property allows the SOM applied to document clustering, to group similar documents together in a cluster and organize similar clusters close together unlike most other clustering methods.

In our proposed approach, we use the implementation of self organizing maps in MATLAB (*Neural Network Toolbox*). We construct a 1-D SOM neural network that takes the generated document vector as input. The size of the network (number of hidden neurons) is based on the desired number of clusters. The network then is trained on the input document vector for about 250 epochs. The output from the network is the weights that define the centers of each cluster. Then we assign each document into its appropriate cluster to be evaluated after that.

Here we list some of the MATLAB-Neural Network Toolbox functions that used in this implementation:

- **newsom**: Create 1-D SOM neural network.
- **train**: Apply SOM training algorithm on input document vectors.
- **sim**: Assign each document vector to its cluster center.

### 3.2.3 Silhouette Coefficient

For clustering, two measures of cluster "goodness" or quality are used. One type of measure allows us to compare different sets of clusters without reference to external knowledge and is called an internal quality measure. The other type of measures lets us evaluate how well the clustering is working by comparing the groups produced by the clustering techniques to known classes which called an external quality measure [7].

In our application of document clustering, we don't have the knowledge of document classes in order to use external quality measures. We will investigate silhouette coefficient (SC Measure) as one of the main internal quality measures.

To measure the similarity between two documents d1 and d2 we use the cosine of the angle between the two document vectors. This measure tries to approach the semantic closeness of documents through the size of the angle between vectors associated to them as in Eq. (2).

$$dist(d_1, d_2) = \frac{d_1 \bullet d_2}{|d_1| \cdot |d_2|} \quad (2)$$

Where $(\bullet)$ denotes vector dot product and (| |) is the dimension of the vector. A cosine measure of 0 means the two documents are unrelated whereas value closed to 1 means that the documents are closely related [18].

Let $\mathbf{D_M} = \{\mathbf{D_1}, .., \mathbf{D_k}\}$ describe a clustering result, i.e. it is an exhaustive partitioning of the set of documents DS. The distance of a document $d \in DS$ to a cluster $\mathbf{D_i} \in \mathbf{D_M}$ is given as in Eq. (3).

$$dist(d, D_i) = \frac{\sum_{p \in D_i} dist(d, p)}{|D_i|} \quad (3)$$

Let further consider $a(d, D_M) = dist(d, D_i)$ being the distance of document d to its cluster Di where $(\mathbf{d} \in \mathbf{D_i})$. $b(d, D_M) = \min_{d \notin D_i} dist(d, D_i) \; \forall D_i \in D_M$ is the distance of document d to the nearest neighbor cluster. The silhouette S (d, DM) of a document d is then defined as in Eq. (4).

$$S(d, D_M) = \frac{b(d, D_M) - a(d, D_M)}{\max(b(d, D_M), a(d, D_M))} \quad (4)$$

The silhouette coefficient (SC Measure) is defined as shown in Eq. (5).

$$SC(D_M) = \frac{\sum_{p \in DS} S(p, D_M)}{|DS|} \quad (5)$$

The silhouette coefficient is a measure for the clustering quality that is rather independent from the number of clusters. Experiences, such as documented in [18], show that values between 0.7 and 1.0 indicate clustering results with excellent separation between clusters, viz. data points are very close to the center of their cluster and remote from the next nearest cluster. For the range from 0.5 to 0.7 one finds that data points are clearly assigned to cluster centers. Values from 0.25 to 0.5 indicate that cluster centers can be found, though there is considerable "noise". Below a value of 0.25 it becomes practically impossible to find significant cluster centers and to definitely assign the majority of data points.

## 4. Experimental Results

The experiments were conducted on three text document datasets EMail1200, SCOTS and Reuters with the three algorithms. There are two main parameters to evaluate the performance of the proposed approach, which are clustering quality and running time.

Document preprocessing step is implemented in Java using NetBeans 5.5.1 and Java API for WordNet Searching (JAWS Library) to access WordNet 2.1 lexical. All the clustering algorithms (k-means, bisecting k-means and SOM neural network) are implemented in MATLAB (Version 7.6.0.324). All experiments were done on Processor P4 (3GHz) machine with 1GB main memory, running the Windows XP Professional® operating system and all times are reported in seconds.

### 4.1 Text Document Datasets

We evaluate the proposed semantic text document clustering approach on three text document datasets: EMail1200, SCOTS and Reuters text corpuses. These datasets vary in the numbers of documents in each dataset, the total number of words, and the average numbers of words in single document. EMail1200 corpus contains test email documents for spam email detection with about 1,245 documents with about 550 words per document. SCOTS corpus (Scottish Corpus Of Text and Speech) contains over 1100 written and spoken texts, with about 4 million words of running text. 80% of this total is made up of written texts and 20% is made up of spoken texts. SCOTS dataset contains about 3,425 words per document. Reuters corpus contains about 21,578 documents that appeared on the Reuters newswire in 1987. The documents were assembled and indexed with categories by

personnel from Reuters Ltd. and Carnegie Group, Inc. in 1987. All the three datasets are used in the text mining testing studies and they are available online for download in [22, 23, 24] respectively.

### 4.2 Clustering Quality

Fig. 2, 3, and 4 show the silhouette coefficient values for the three datasets respectively. In all experiments SOM neural network outperforms k-means and bisecting k-means algorithms in the overall clustering quality using silhouette measure.

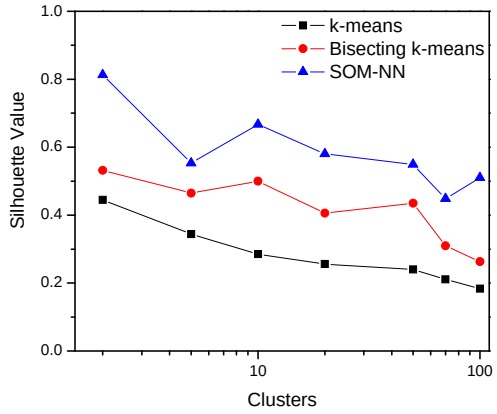



Fig. 2 Silhouette values comparing all clustering algorithms – EMail1200 Dataset

From these figures we notice the good clustering quality results obtained by SOM-NN with comparison to other algorithms. For example, at number of clusters (k = 2), we found that SC value for SOM for EMail1200 dataset is about 0.813 which considered an excellent clustering result with well separated clusters. If we check the other algorithms results, we found that bisecting k-means overcomes basic k-means algorithm with SC value equal to 0.532 which means that the data points are clearly assigned to cluster centers.

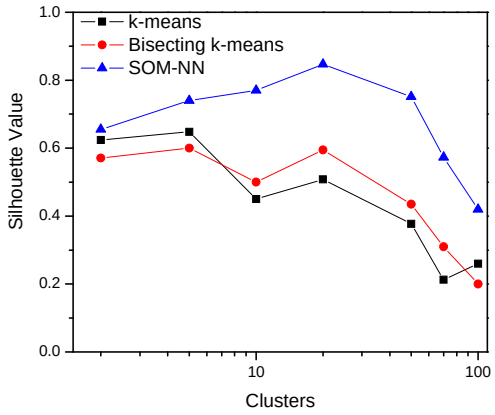



Fig. 3 Silhouette values comparing all clustering algorithms – SCOTS Dataset

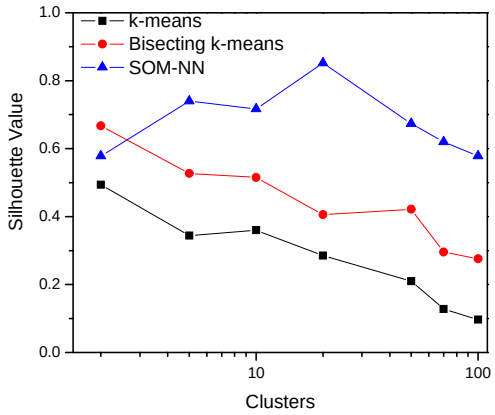



Fig. 4 Silhouette values comparing all clustering algorithms – Reuters Dataset

For SCOTS dataset, as in fig. 3, we found that k-means and bisecting k-means algorithms nearly generates the same clusters. However SOM outperforms other algorithms at k=20. SOM-NN achieves silhouette value equal to 0.847 where other algorithms obtain about 0.595 and 0.508 respectively. The last experiment results in fig. 4 show the great performance optimization between SOM-NN and other algorithms in Reuters datasets.

We have performed two more experiments to show the effect of using WordNet lexical categories with SOM neural network on the final clustering quality results. We measure SC value for SOM on both SCOTS and Reuters datasets in two cases: first using traditional bag-of-words technique, second using WordNet lexical categories. Fig. 5 and 6 show the silhouette coefficient values for the two datasets.

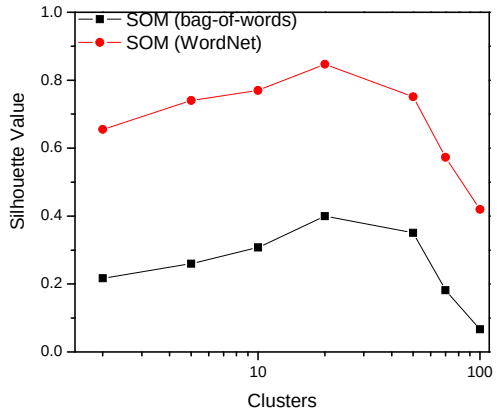



Fig. 5 WordNet improves SOM clustering results using SCOTS dataset

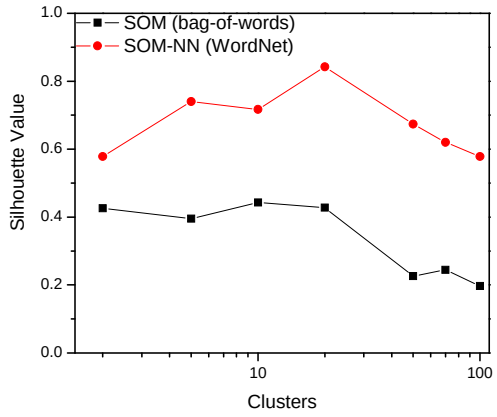



Fig. 6 WordNet improves SOM clustering results using Reuters dataset

For SCOTS dataset the clustering results is very good, because the proposed approach overcomes the traditional approach with about 3 times. The clustering results for Reuters dataset is also positive. The proposed approach achieves about twice clustering quality than traditional technique. This experiment shows that using WordNet lexical categories in the feature extraction process improves the overall clustering quality of the input dataset document than traditional approaches that uses bag-of-words technique.

### 4.3 Running Time

Reuters dataset, as mentioned early in this section, contains about 21,578 documents. This is considered a real challenge task that faces any clustering approach because of "Scalability". Some clustering techniques that are helpful for small data sets can be overwhelmed by large data sets to the point that they are no longer helpful. For that reason we test the scalability of our proposed approach with the different algorithms using Reuters dataset. This experiment shows that the SOM neural network performs a great running time optimization with comparison to other two algorithms. Also, according to the huge size of Reuters dataset, the proposed approach shows very good scalability against document size.

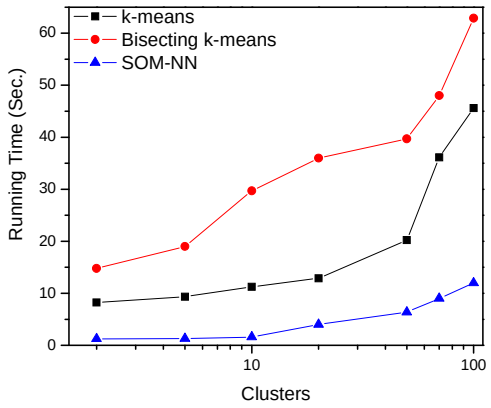



Fig. 7 Scalability of all clustering algorithms on Reuters dataset

Fig. 7 depicts the running time of the different clustering algorithms using Reuters dataset with respect to different values of desired clusters. The overall process of document clustering using WordNet lexical categories is done in a very low time in comparison with other two approaches. SOM neural network achieves speed-up ratio 10 times faster than bisecting k-means algorithm and about 5 times faster than basic k-means algorithm for Reuters dataset.

## 5. Conclusion

In this paper we proposed a semantic text document clustering approach based on the WordNet lexical categories and SOM neural network. The proposed approach generates documents vectors using the lexical category mapping of WordNet after preprocessing the input documents. We apply three different clustering algorithms, SOM neural network, k-means, and bisecting k-means to the generated documents vectors. The output clusters in each case are evaluated using silhouette coefficient measure to test the performance of the proposed approach. The results show that SOM neural network achieves higher clustering quality than other two clustering algorithms k-means, and bisecting k-means. Also, the results show that by using WordNet lexical categories in the feature extraction process for text documents improves the overall clustering quality. Finally, the proposed approach shows good scalability against the huge number of documents as in Reuters dataset along with different values of desired clusters.

## References

[1] R. Saraçoğlu, K. Tütüncü, and N. Allahverdi, (2007) "A fuzzy clustering approach for finding similar documents using a novel similarity measure," Expert Systems with Applications, vol. 33, no. 3, pp. 600–605.

[2] H. X. W. Wu and S. Shekhar, (2003)Eds., Clustering and Information Retrieval. Kluwer.

[3] K. Bade and A. Nurnberger,( 2006) "Personalized hierarchical clustering," in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence. Washington, DC, USA: IEEE Computer Society, pp. 181–187.

[4] Z. Jiang, A. Joshi, R. Krishnapuram, and L. Yi, (2000) "Retriever: Improving web search engine results using clustering," University of Maryland Baltimore County, Technical Report.

[5] J. Sedding and D. Kazakov, (2004) "WordNet-based Text Document Clustering", COLING 3rd Workshop on Robust Methods in Analysis of Natural Language Data, pp. 104–113, Geneva, Switzerland.

[6] M. Lan, C.L. Tan, H.B. Low and S.Y. Sung, (2005) "A Comprehensive Comparative Study on Term Weighting Schemes", Proceedings of the 14th International World Wide Web (WWW2005) Conference, Japan, pp.1032–1033.

[7] B.B. Wang, R.I. McKay, H.A. Abbass and M. Barlow, (2002) "Learning text classifier using the domain concept hierarchy", In Proceedings of International Conference on Communications, Circuits and Systems, China, pp. 1230–1234.

[8] W.K. Gad and M.S. Kamel, (2009) "Enhancing text clustering performance using semantic similarity", Lecture Notes in Business Information Processing, 24 LNBIP, pp. 325-335.

[9] L. Jing, M.K. Ng and J.Z. Huang, (2010) "Knowledge-based vector space model for text clustering", Knowledge and Information Systems, 25 (1), pp. 35-55.

[10] O. Zamir, O. Etzioni, O. Madani, and R.M. Karp, (1997)"Fast and intuitive clustering of web documents", In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, KDD97, pp. 287–290.

[11] M. Steinbach, G. Karypis and V. Kumar, (2000)"A Comparison of Document Clustering Techniques", Department of Computer Science and Engineering, University of Minnesota, Technical Report , #00-034.

[12] F. Beil, M. Ester and X. Xu, (2002) "Frequent term-based text clustering", Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD02), Edmonton, Alberta, Canada, pp. 436–442.

[13] S.J. Green, (1999) "Building hypertext links by computing semantic similarity", IEEE Transactions on Knowledge and Data Engineering, Vol.11, pp.713–730.

[14] S.J. Green, (1997) "Building hypertext links in newspaper articles using semantic similarity", The 3rd Workshop on Applications of Natural Language to Information Systems, NLDB 97, pp. 178–190.

[15] D.M.P.K. Dave and S. Lawrence, (2003) "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of the 12th International World Wide Web Conference, Budapest, Hungary, pp. 519–528.

[16] A. Hotho, S. Staab and G. Stumme, (2003) "Wordnet improves text document clustering", ACM SIGIR 2003 Workshop on Semantic Web, pp. 541–544.

[17] D. Reforgiato, (2007) "A new unsupervised method for document clustering by using WordNet lexical and conceptual relations", Journal of Information Retrieval, Vol. 10, pp.563–579.

[18] M. Friedman. A. Kandel, M. Schneider, M. Last, B. Shapka, Y. Eloviciand O. Zaafrany, (2004) "A Fuzzy-Based Algorithm for Web Document Clustering. Fuzzy Information", Processing NAFIPS '04, IEEE Annual Meeting of the North American, Vol. 2, pp. 524–527.

[19] C. Hung and S. Wermter, (2004) "Neural Network-based Document Clustering Using WordNet Ontologies", International Journal of Hybrid Intelligent Systems, Vol. 1, pp. 127-142.

[20] U.K. Sridevi and N. Nagaveni (2011) "Semantically Enhanced Document Clustering Based on PSO Algorithm", European Journal of Scientific Research, Vol.57, No.3, pp. 485-493.

[21] L. Kaufman and P.J. Rousseeuw, (1999) "Finding Groups in Data: an Introduction to Cluster Analysis", Published by John Wiley & Sons, USA.

[22] EMail1200 dataset: http://boole.cs.iastate.edu/book/acad/bag/data/lingspam

[23] SCOTS dataset: http://www.scottishcorpus.ac.uk/

[24] Reuters dataset: http://www.daviddlewis.com/resources/testcollections/reuters21578/

# Comparison Based Analysis of Different Cryptographic and Encryption Techniques Using Message Authentication Code (MAC) in Wireless Sensor Networks (WSN)

**Sadaqat Ur Rehman*, Muhammad Bilal**, Basharat Ahmad**, Khawaja Muhammad Yahya**, Anees Ullah**, Obaid Ur Rehman***

***Department of Electrical Engineering, Sarhad University of Science and Technology, Peshawar, 25000, Pakistan.**
****Department of Computer Systems Engineering, N-W.F.P. University of Engineering & Technology Peshawar, 25000, Pakistan.**

**Abstract**

Wireless Sensor Networks (WSN) are becoming popular day by day, however one of the main issue in WSN is its limited resources. We have to look to the resources to create Message Authentication Code (MAC) keeping in mind the feasibility of technique used for the sensor network at hand. This research work investigates different cryptographic techniques such as symmetric key cryptography and asymmetric key cryptography. Furthermore, it compares different encryption techniques such as stream cipher (RC4), block cipher (RC2, RC5, RC6 etc) and hashing techniques (MD2, MD4, MD5, SHA, SHA1 etc). The result of our work provides efficient techniques for communicating device, by selecting different comparison matrices i.e. energy consumption, processing time, memory and expenses that satisfies both the security and restricted resources in WSN environment to create MAC.

***Keywords***: MAC, WSN, parameter, cryptographic techniques, stream cipher, block cipher, hashing techniques.

## 1. Introduction

Wireless sensor networks (WSN) have the advantage over traditional networks in many ways such as large scale, autonomous nature and dense deployment [1]. Moreover, it has increased fault tolerance because if a sensor node fails others can collect/process data. Because of its ad-hoc nature it becomes more attractive in certain applications such as military, environmental observation, syndrome surveillance, supply chain management, fire detection, vision enabling, energy automation, building administration, gaming, health and other commercial and home applications [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15].

With the wide deployment of WSN for multi-faceted applications security is becoming a growing concern. For example, in a battlefield, a military communication network used for sensitive information interchange can be hacked by its adversaries if the WSN has security holes causing severe loss of life and machinery. Similarly, many social problems can be created if personal information flowing on health care systems is intercepted [16]. Security of WSN is a big challenge due to its limited resources such as energy, power supplies, small memory, computation and communication capabilities [17], [18], [19], [20], [1]. This is the reason that traditional security techniques cannot be applied on sensor networks, indirectly rising the need to make sensor network economically feasible [21], [22].

Cryptographic algorithm plays an important role in the security and resource conservation of wireless sensor networks (WSN) [23], [24]. This work spotlights different cryptographic techniques and compares different encryption techniques such as stream cipher (RC4), block cipher (RC2, RC5, RC6 etc) and hashing techniques (MD2, MD4, MD5, SHA, SHA1 etc). Our main aim of working in this survey paper is to put forward a cryptographic and encryption technique that creates Message Authentication Code (MAC) in wireless sensor networks (WSN), which is more feasible in the restricted resources of wireless sensor networks (WSN) and also provide good security in communication as well.

The rest of the paper is outlined as follows. We present our critical review questions in section II; then we analyze these review questions in section III and finally we deduct our conclusion in section IV.

## 2. Research Questions

The critical review questions we are in quest of to answer are:

- Why we prefer symmetric keys over public keys in WSN?
- Which method and algorithm is best to create MAC in Wireless Sensor Networks?

- Can we apply all given methods to create MAC like block cipher, stream cipher, hash function and unconditional secure?

## 3. Analysis and Discussion

In this section we analyze the result of our research questions. Our questions are:

### 3.1 Why we prefer symmetric keys over public keys in WSN?

Symmetric key technique uses a single key called secret key which uses less mathematics, results in less computation, on the other hand asymmetric key technique uses both public and private keys, results in more processing and consumes more energy. Symmetric key techniques offer better energy efficiency as compared to public key that is why most researches use it for creating MAC in WSN.

According to [25] public key is used in some applications for secure communications e.g. SSL (Secure Socket Layer) and IPsec standards both use it for their key agreement protocols. But it consumes more energy and also it is more expensive as compared to symmetric key.

[26] has given a reason that public key consumes more energy due to great deal of computation and processing involved, which makes it more energy consumptive as compared to symmetric key technique e.g. a single public key operation can consume same amount of time and energy as encrypting tens of megabits using a secret key cipher.

According to [27], the more consumption of computational resources of public key techniques is due to the fact that it uses two keys. One of which is public and is used for encryption, and every one can encrypt a message with it and other is private on which only decryption takes place and both keys has a mathematical link, the private key can be derived from a public key. In order to protect it from attacker the derivation of private key from the public is made difficult as possible like taking factor of a large number which makes it impossible computationally. Hence, it shows that more computation is involved in asymmetric key technique thus we can say that symmetric key is better to choose for WSN.

According to [28] the cost of public key is much more expensive as compared to symmetric key for instance, *a 64 bit RC5 encryption on ATmega128 8MHz takes 5.6 milliseconds, and a 160 bit SHA1 has function evaluation takes only 7.2 millisecond's.* These symmetric key algorithms are more than 200 times faster than Public key algorithms.

Public key cryptography is not only expensive in computation but also it is more expensive in communication as compared to symmetric key cryptography. According to [4] to send a public key from one node to another, at least 1024 bits required to be sent if the private key is 1024 bits long.

[25], [26], [27] and [28] suggest that symmetric key cryptography is better than asymmetric key cryptography in both cost and computation.

### 3.2 Which method and algorithm is best to create MAC in Wireless Sensor Networks?

According to [29] Block Cipher is more secure as compared to Stream Cipher this is because of the facts:

- Attacks such as differential attacks on block cipher are also applied to stream cipher.
- Attacks such as correlation attacks on stream cipher are not valid on block cipher.
- Algebraic attacks on stream cipher are more effective.
- Guess and set attacks against stream ciphers recover the key or any plaintext.
- Generic time/memory attacks are stronger against stream cipher than block cipher.

Because of these facts it shows that block cipher is more secure as compared to stream cipher and thus the stream cipher will be replaced with block cipher except few applications.

[30] Compared different attacks on Hash function like birthday attack. He uses MD5 algorithm for this attack and finds out that such an attack needs $2^{64}$ blocks (or $2^{73}$ bits) of data for authentication using the same key. If the communication link has the ability to process 1 Gbit/sec it means one need 250,000 years to process the data needed by such an attack. Even according to [30] on software implementation the popular hash function is faster than the block cipher.

As stream cipher uses a key "K" and initialization vector (IV) for encryption making it more vulnerable to retrieve the plaintext in case different packets use the same IV. If initialization vector is long then it will require additional bytes but our aim is to reduce the packet overhead. Thus we follow the principle *"use an encryption scheme that is as robust as possible in the presence of repeated IVs"*. As stream cipher does not follow this principle so the only way is to use block cipher.

Block cipher has different algorithms such as DES, AES, RC5 and Skipjack. The block cipher used for encryption has an extra advantage i.e. *the most efficient MAC algorithm use a block cipher* [31].

After choosing block cipher for creating MAC, we need to choose algorithm in block cipher. DES is very slow when it is implemented in software. Similarly, experimentations show that AES is quite slow. We find that RC5 and Skipjack are more suitable for sensor networks. One can outperform the other on specific hardware platform. For example, on TinySec platform, although RC5 is slightly faster than Skipjack but it

uses 104 extra bytes of RAM per key for good performance. Therefore the default block cipher in TinySec is Skipjack [31]. [31] tested the performance of RC5 and Skipjack on Mica2 sensor node to determine the speed of these two 64 bit block cipher. The time to execute cipher operation on the Mica 2 sensor node is shown as in the Table 1.

Table 1: Time to execute cipher operation on the mica2 sensor nodes [31]

| ***Block Cipher*** | ***Time (ms)*** | ***Time (byte times)*** |
|---|---|---|
| RC5 ( C ) | 0.90 | 2.2 |
| Skipjack ( C ) | 0.38 | 0.9 |
| RC5 ( C, assembly) | 0.26 | 0.6 |

[28] has chosen five popular encryption schemes for study which ranges from stream cipher (RC4) and block ciphers (RC5, IDEA) to hashing techniques ( SHA-1, MD5). RC5 was also chosen for Atmega in the Berkeley Motes SPINS Project. RC5 was chosen on this platform because it uses less memory. [28] Also found that hashing techniques requires an order of magnitude higher overhead.

The parameters used in our paper are shown as in Table 2.

Table 2: Encryption schemes and parameters [28]

| ***Algorithm*** | ***Type*** | ***Key/Hash*** | ***Block*** |
|---|---|---|---|
| RC4 [2] | Stream | 128 bits | 8 bits |
| IDEA [2] | Block | 128 bits | 64 bits |
| RC5 [1] | Block | 64 bits | 64 bits |
| MD5 [2][3] | 1-way hash | 128 bits | 512 bits |
| SHA1 [4] | 1-way hash | 128 bits | 512 bits |

On hardware platforms [28] evaluates the performance of these different cryptographic algorithms on different processors that ranges from low end i.e. (4 MHz 8 bit Atmel AVR Atmega 103) to high end (400 MHz 32 bit Intel XScale). Which are shown as in Table 3.

Table 3: Hardware platforms [28]

| ***Platform*** | ***Word Size*** | ***Clock Frequency*** | ***I/D-S*** |
|---|---|---|---|
| Atmega 103 | 8 bits | 4 MHz | None |
| Atmega 128 | 8 bits | 16 MHz | None |
| M16C/10 | 16 bits | 16 MHz | None |
| SA-1110 | 32 bits | 206 MHz | 16/8 KB |
| PXA250 | 32 bits | 400 MHz | 32/32 KB |
| UltraSparc2 | 64/32 bits | 440 MHz | 16/16 KB |

Experiments have been performed for different values of selected parameters on all these algorithms, architecture and considered platforms. The functional block of all these algorithms i.e. initialization, encryption and decryption was executed 1000 times using the same input and the result was averaged for these execution.

The execution time overhead for each algorithm and for considered platforms on a log scale is shown as in the Figure1. These are also shown in Table 4.

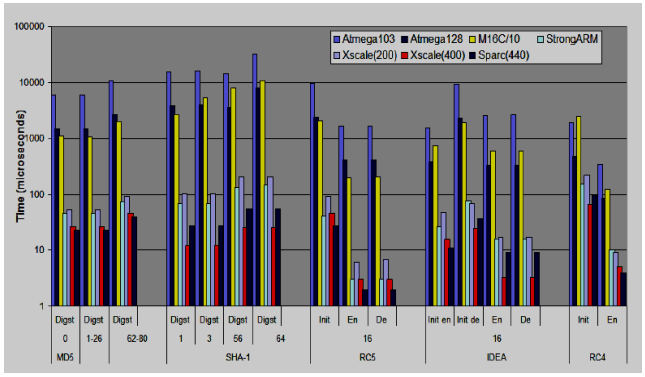



Fig. 1 Execution times [µs] for algorithms, platforms and plaintext sizes [bytes] [28]

Table 4: Execution times [µS] for algorithms, platforms and plaintext sizes [bytes] [28].

| Algorithm | Size | Action | Atmega103 | Atmega128 | M16C/10 | StrongARM | Xscale(400) | Xscale(200) | Sparc(440) |
|---|---|---|---|---|---|---|---|---|---|
| MD5 | 0 | Digest | 5863 | 1466 | 1083 | 46 | 26 | 53 | 23 |
| | 1-26 | Digest | 5890 | 1473 | 1075 | 46 | 26 | 53 | 23 |
| | 62-80 | Digest | 10888 | 2722 | 2011 | 74 | 45 | 90 | 39 |
| | | | | | | | | | |
| SHA-1 | 1 | Digest | 15249 | 3812 | 2651 | 69 | 12 | 102 | 27 |
| | 3 | Digest | 15781 | 3945 | 5303 | 69 | 12.3 | 103 | 27 |
| | 56 | Digest | 14543 | 3636 | 7955 | 133 | 25.8 | 205 | 55 |
| | 64 | Digest | 31107 | 7777 | 10907 | 145 | 25.7 | 207 | 56 |
| | | | | | | | | | |
| RC5 | 16 | Init | 9641 | 2410 | 2074 | 41 | 45 | 91 | 28 |
| | | Enc | 1651 | 413 | 197 | 3 | 3 | 6 | 2 |
| | | Dec | 1636 | 409 | 202 | 3 | 3 | 7 | 2 |
| | | | | | | | | | |
| IDEA | 16 | Init enc | 1523 | 381 | 727 | 26 | 15.54 | 47 | 11 |
| | | Init dec | 9417 | 2354 | 1927 | 76 | 25.16 | 69 | 36 |
| | | Enc | 2555 | 325 | 596 | 16 | 3.24 | 17 | 9 |
| | | Dec | 2614 | 325 | 597 | 16 | 3.27 | 17 | 9 |
| | | | | | | | | | |
| RC4 | | Init | 1886 | 472 | 2455 | 155 | 66.8 | 216 | 96 |
| | | Enc | 344 | 86 | 123 | 10 | 5 | 9 | 4 |

After performing simulation of these algorithms [28] summarizes the result in table V. Comparing the RC4 and RC5 on Atmega 103 shows that the encryption time for both algorithms are close to each other, in fact, RC4 is slightly faster. But, however, by comparing them on Strong ARM, it shows that RC5 is 3 times faster than RC4 algorithm although RC4 operates on 8 bits while RC5 operates on 32 bits.

Comparing RC5 with IDEA on the Atmega 103 showed that RC5 is 1.5 times faster than IDEA. However, both of these algorithms use 64 bit blocks. Hashing techniques needs almost an order of a magnitude higher overhead. Thus RC5 is faster compared to other algorithms like RC4 and IDEA so it requires less processing time and thus less energy consumption.

Table 5: Encryption algorithm memory usage on micaz and telosb sensor motes [32]

| ***Encryption Algorithm*** | ***MicaZ*** | | ***TelosB*** | |
|---|---|---|---|---|
| | ***RAM (KB)*** | ***ROM (KB)*** | ***RAM (KB)*** | ***ROM (KB)*** |
| RC5 | 0.2 | 2.5 | 0.2 | 6 |
| AES | 2 | 10 | 1.8 | 9 |
| Skipjack | 0.6 | 10 | 0.04 | 7.5 |
| XXTEA | 0.049 | 3.1 | 0.04 | 3.8 |

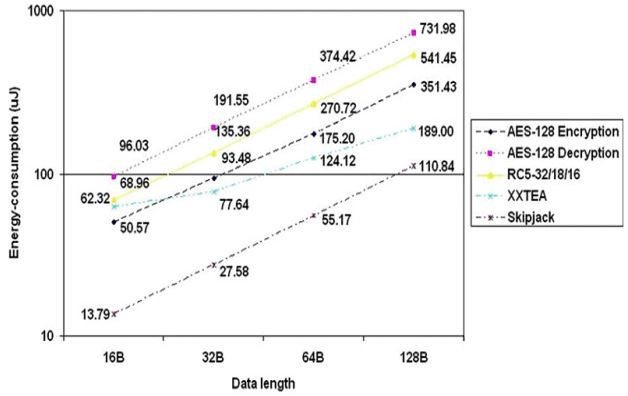



Fig. 2 Energy consumption of block cipher on Micaz sensor motes [32].

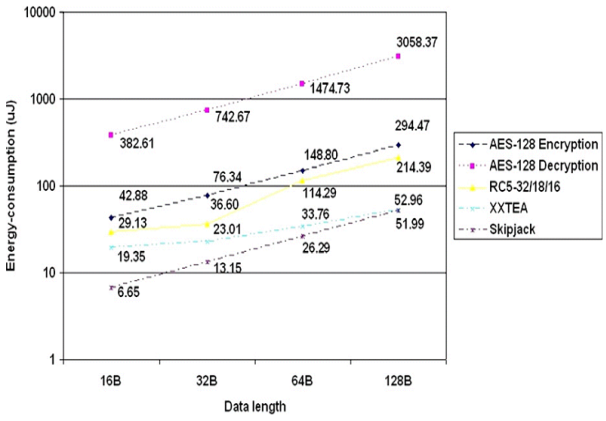



Fig. 3 Energy consumption of block cipher on TelosB sensor motes [32].

[32] Shows Skipjack and XXTEA have efficient energy consumption and smaller memory requirements while RC5 and AES provide better security. The energy consumption on RC5 and AES depends on the key size and number of rounds respectively. RC5 consumes more energy in its encryption phase than that of AES on MicaZ, but overall energy consumption and memory of RC5 is less. Similarly on TelosB, RC5 consume less energy for both encryption and decryption than AES and uses less memory.

RC5 provide good security against different attacks and is also a fast block cipher algorithm suitable for both hardware and software implementation. Since it is a parameterized algorithm having variable features (like block size, number of rounds and length secret key) it provides flexibility in both performance and security [31], [33].

### 3.3 Can we apply all given methods to create MAC like block cipher, hash function and stream cipher etc?

Yes we can, but we have to look to the resources and we need to choose such a technique which is feasible for sensor networks. In case of more resources the best option is to go for hash function because it provides better security. But according to the current situation, sensor network has limited resources, it is common practice to use block cipher for implementation of MAC in sensor networks as it requires less resources comparatively.

Thus, from the above discussion we conclude that block cipher is the best option to create MAC in WSNs.

## 4. Conclusion

In this paper we investigate that public key is not energy efficient and is expensive in terms of both computation and communication as compared to symmetric key. Sensor networks has limited resources, therefore most of the researcher used symmetric key to create MAC in WSNs. Thus, we conclude that symmetric key techniques are more feasible for WSNs as compared to public key.

After selecting symmetric key techniques, we compared different attacks on hashing techniques and conclude that it offers good security mechanisms as compared to block cipher. However, it requires an order of magnitude higher overhead and also uses more memory. While stream cipher, Skipjack and XXTEA are less secure than block cipher and for encryption packets overhead also takes place in stream cipher. By selecting an efficient technique, we pick block cipher as best technique to create Message authentication code (MAC) in sensor network although hash function offers good security.

We conclude that RC5 is feasible and consumes less energy/resources as compared to other algorithms (AES, MD5, SHA1, IDEA) except Skipjack and XXTEA. However, RC5 is more secure than Skipjack and XXTEA. Thus we propose that RC5 is a best algorithm to create MAC in sensor networks.

## References

[1] Matthew N. Vella, Texas A&M University-Corpus Christi, Computer Science Program, Dr. Ahmed Mahdy Texas A&M University-Corpus Christi, Computer Science Faculty "Survey of Wireless Sensor Network Security"

[2] Chung-Kuo Chang, J. Marc Overhage, Jeffrey Huang "An Application of Sensor Networks for Syndromic Surveillance" 2005 IEEE

[3] Dunfan Ye,DaoliGong,WeiWang "Application of Wireless Sensor Networks in Environmental Monitoring", 2009 2nd International Conference on Power Electronics and Intelligent Transportation System.

[4] Ling Tan, Shunyi Zhang, and Yanfeng Sun, Jing Qi "Application of Wireless Sensor Networks in Energy Automation", Sustainable Power Generation and Supply, 2009. Supergen '09. International conference

[5] Sundip Misra, Vivek Tiwari and Mohammad S. Obaidat, *Fellow*, IEEE "LACAS: Learning Automata-Based Congestion Avoidance Scheme for Healthcare Wireless Sensor Networks", IEEE Journal on Selected Areas in Communications, Vol. 27, No. 4, May 2009

[6] Ian F. Akylidiz, Fellow IEEE, Tommaso Melodia, Member IEEE, and Kaushik R. Chowdhury, Student Member IEEE "Wireless Multimedia Sensor Networks: Applications and Testbeds", Proceedings of the IEEE. Vol. 96, No. 10, October 2008

[7] Kwangsoo Kim, Jongarm Jun, Sunjoong Kim, and Byung Y. Sung "Medical Asset Tracking Application in Wireless Sensor Networks", The Second International Conference on Sensor Technologies and Applications, 2008 IEEE

[8] N. Rajendran, P. Kamal, D. Nayak, and S. A. Rabara, " WATS-SN: A Wireless Asset Tracking System using Sensor Networks", Proceedings of IEEE International Conference On Personal Wireless Communications, Jan 2005

[9] G. W. Allen, K. Lorinca, M. Welsh, O. Marcillo, J. Johnson, M. Ruiz, and J. Lees, "Deploying a Wireless Sensor Network on an Active Volcano", IEEE Internet Computing, IEEE Computer society, March/April 2006

[10] K. Chintalapudi, T. Fu, J. Paek, N. Kothari, S. Rangwala, J.Caffrey, R. Govindan, E. Johnson, "Monitoring Civil Structures with a Wireless Sensor Network", IEEE Internet Computing, IEEE Computer society, March/April 2006

[11] I. Ituen and G. Sohn, "The Environmental Applications of Wireless Sensor Networks", International Journal of Contents, Vol.3, No. 4, Dec 2007

[12] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, J. Anderson, "Wireless Sensor Networks for Habitat Monitoring", WSNA'02, Sep 2002

[13] Anthony Rowe, Dhiraj Goel, Raj Rajkumar "FireFly Mosaic: A Vision-Enabled Wireless Sensor Networking System", 28th IEEE International Real-Time Systems Symposium. 2007 IEEE

[14] E. Sazonov, K. Janoyan, and R. Jha, "Wireless Intelligent Sensor Network for Autonomous Structural Health Monitoring", Proceedings of Structural Materials Technology (SMT): NDE/NDT for Highways and Bridges, 2004

[15] http://corporate.traffic.com

[16] Chien-Wen Chiang, Chih-Chung Lin, Ray-I Chang "A New Scheme of Key Distribution using Implicit Security in Wireless Sensor Networks" Feb. 7-10, 2010 ICACT 2010.

[17] Xiaojiang Du, North Dakota State University and Hsiao-Hwa Chen, National Cheng Kung University "Security in Wireless Sensor Networks" IEEE Wireless Communication August 2008

[18] Sung-Chul Jung , Hyoung-Kee Choi. School of Information and Communication Engineering "An Energy-aware Routing Protocol Considering Link-Layer Security in Wireless Sensor Networks." Feb.15-18, 2009 ICACT 2009

[19] Md. Anisur Rahman and Mitu Kumar Debnath "An Energy-Efficient Data Security System for Wireless Sensor Network" Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008) 25-27 December, 2008, Khulna, Bangladesh

[20] Al-Sakib Khan Pathan, Hyung-Woo Lee, Choong Seon Hong "Security in Wireless Sensor Networks: Issues and Challenges" Feb. 20-22, 2006 ICACT 2006

[21] Perrig A, Stankovic J and Wagner D (2004) "*Security in Wireless Sensor Networks*", In Communication of the ACM, 47(6), June 2004.

[22] Ali Nur Mohammad Noman "A Generic Framework For Defining Security Environments Of Wireless Sensor Networks" 5th International Conference on Electrical and Computer Engineering ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

[23] Mohammad AL-Rousan, A.Rjoub and Ahmad Baset "*A low-energy security algorithm for exchanging information in wireless sensor networks*", Journal of information assurance and security 4 (2009) 48-59.

[24] Y.W. Law, S. Dulman, S. Etalle, P. Havinga (2002), "*Assessing security critical energy efficient sensor network*", Available at: http://www.dsv.su.se/~matei/bin/4%20-%202i1279/L5_EYES.pdf

[25] Ning P, Wang R and Du W (2005), "*An efficient scheme for authenticating public keys in sensor networks*", Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, Chicago, IL, USA, pp. 58-67.

[26] Goodman J and Chandrakasan P (2001), "*An Energy Efficient Reconfigurable Public Key Cryptography Processor*", IEEE journal of solid state circuits, pp. 1808-1820, November 2001.

[27] RSA Security (2004), "*Cryptography*", Available at: http://www.rsasecurity.com/rsalabs/node.asp?id=2152,

[28] Ganesan P, Venugopalan R, Peddabachagari P, Dean A, Mueller F and Sichitiu M (2003), "*Analyzing and modelling encryption overhead for sensor network nodes*", In Proceeding of the Ist ACM international workshop on Wireless sensor networks and application, San Diego, California, USA, September 2003.

[29] Dasgupta A (2005), "*Analysis of Different Types of Attacks on Stream Ciphers and Evaluation and Security of Stream Ciphers*", Available at: http://www.securitydocs.com/library/3235,

[30] Bellare M, Canetti R and Krawczyk H (1996), "*Message Authentication using Hash Functions The HMAC Construction*", Appears in RSA Laboratories CryptoBytes, Vol. 2, No. 1, Spring 1996.

[31] Karlof C, Sastry N and Wagner D (2004), "*TinySec: A Link Layer Security Architecture for Wireless Sensor Networks*", Proceedings of the second ACM Conference on Embedded Networked Sensor Systems, November 2004

[32] Jongdeog Lee, krasimira kapitanova, sang H. son "The price of security in wireless sensor networks" computer networks 54 (2010) 2967-2978 available at www.elsevier.com/locate/comnet

[33] Afrin Zahra, M. Nizam uddin & Z.A Jaffery "Implementation and Analysis of Security Protocols for Wireless Sensor Network" International journal of Electronics Engineering, 2(1), 2010, pp. 111-113.

**First Author** Sadaqat Ur Rehman has completed his BS degree program in Computer System Engineering from N.W.F.P University of Engineering and Technology Peshawar, his areas of interest are Wireless sensor Networks, Artificial Intelligence and Microcontroller. Currently he is Lab Engineer in

Electrical Engineering Department at Sarhad University of Sciences and Information Technology Peshawar

**Second Author** Muhammad Bilal is a student of BS degree program in Computer System Engineering at N.W.F.P University of Engineering and Technology Peshawar, his areas of interest are Wireless sensor Networks, Digital Image Processing and Artificial Intelligence.

**Third Author** Basharat Ahmad is a student of BS degree program in Computer System Engineering at N.W.F.P University of Engineering and Technology Peshawar, his areas of interest are Wireless sensor Networks, Microcontroller and Data Base Management System.

**Fourth Author** Khawaja Muhammad Yahya has completed his MS degree in Computer Engineering from University of Missouri-Rolla, USA in 1987. He Completed PhD in Information Management System / Decision Support System from University of Missouri-Rolla, USA in 1995. Currently he is a Chairman of Department of computer System Engineering at N.W.F.P University of Engineering and Technology Peshawar.

**Fifth Author** Anees Ullah has completed his BS degree in Electrical Engineering from N.W.F.P University of Engineering and Technology Peshawar; currently he is pursuing MS degree from the same university.

**Sixth Author** Obaid Ur Rehman has completed his BS degree in Electrical Engineering from N.W.F.P University of Engineering and Technology Peshawar, MS degree from university of Liverpool UK. Currently he is Assistant professor in Electrical Engineering Department at Sarhad University of Sciences and Information Technology Peshawar.

# Fuzzy-controlled Load-balanced Broadcasting (FLB) In Clustered Mobile Ad Hoc Networks

**Anuradha Banerjee**

**Kalyani Govt. Engg. College**
**Kalyani, Nadia, West Bengal, India**

## Abstract

**Problem statement**: In mobile ad hoc networks owing to node mobility, broadcasting is expected to be more frequently used to find route to a particular destination, to page a host and to alarm all hosts. The simplest and commonly used mechanism for broadcasting is flooding, where every node retransmits every uniquely received message exactly once. Despite its simplicity, it can result in highly redundant retransmission, contention and collision in the network, a phenomenon referred to as broadcast storm problem. Several approaches have been proposed to mitigate this problem inherent with flooding. However, none of those schemes guarantees minimum redundancy with 100% delivery ratio. **Present Approach:** The present study proposes a fuzzy-controlled load-balanced broadcast scheme (FLB) in a multi-hop clustered ad hoc network that guarantees complete packet delivery at no redundancy. Each node $n_j$ elects its most eligible uplink neighbor $n_i$ within its cluster and that uplink neighbor $n_i$ has to take the responsibility of transmitting all broadcast messages to $n_j$. Hence, the redundancy is zero. **Results:** Simulation results show that the proposed broadcast algorithm provides high packet delivery ratio at minimum overhead and minimum delay, w.r.t. other state-of-the-art broadcast algorithms.

***Keywords: Ad hoc network, Broadcasting, Fuzzy, Load-balance, Redundancy.***

## 1. Introduction

A mobile ad hoc network is a wireless network that is self-organized with many mobile nodes. No static infrastructure such as a wired backbone is available. All nodes are free to move around and the network topology may change frequently. Due to limited transmission range of wireless network interface, nodes are required to forward messages for those located outside the radio-coverage, thereby forming a multi-hop network. Possible applications include emergency rescue in disaster situations, communication between mobile robots, exchanging information in the battlefield etc. [1-5]. Each node can directly send information in single hop within a pre-specified circle around the node. That circle is called radio-circle and its radius is called radio-range. If a node $n_j$ stays within the radio-circle of another node $n_i$ at time t, then $n_j$ will be called a downlink neighbor of $n_i$ at time t and $n_i$ will be called an uplink neighbor of $n_j$ at that time. In this situation $n_i$ can directly transmit information to $n_j$ without the assistance of any intermediate node as router. Otherwise, the communication between the nodes $n_i$ and $n_j$ is multi-hop.

Broadcast is a common operation in ad hoc networks. By broadcast, a message is propagated to all nodes in the network. The problem of redundancy is highly involved in case of broadcasting. For example, if a node has multiple uplink neighbors, then it will receive the broadcast message from all those uplink neighbors resulting in redundancy.

Broadcast is useful in delivering messages to users with unknown location or group of users whom the source need not exactly know [5]. Broadcast plays an important role in routing, network management etc. Many on-demand or reactive routing protocols (dynamic source routing (DSR) [2], ad hoc on-demand distance vector routing (AODV) [3], on-demand multicast routing protocol (ODMRP) [4] etc.) rely on broadcast to discover a route between two nodes or to update group status and multicast routes. Broadcast is also a viable candidate for multicast in ad hoc networks with rapid changing topology. In the next section I discuss some state-of-the-art broadcast algorithms.

## 2. Related Work

A density based innovative flooding (DBF) algorithm is proposed in [6]. In this algorithm, each node forwards a message based on its neighbor density and neighbor density of its previous node from which the broadcasted message. In a cluster of loosely couples nodes with few intermediate nodes as neighbors, the probability of forwarding the broadcasted message will be high. On the other hand, if a node is having high density of neighbors, then there will be lots of chances of packet collision at that

point. Density based flooding tries to avoid that situation by assigning low priority at that point.

The article in [7] proposes a tree based broadcast (TBB) method that maintains a spanning tree in the network. The algorithm is fully distributed, decentralized and resource-efficient. Broadcast operation is performed using a tree by forwarding the message not to all neighbors, but only those neighbors in the tree structure. Since the tree is acyclic, each message is received only once by each node, giving two advantages over the existing methods. Firstly it is needless to store the previous broadcasts in order to avoid endless multiplications of broadcast messages along a cycle of links. Only the originator of a broadcast message needs to store it and pay attention to whether its broadcast was successful or not if it is of great importance. Secondly, it is very economical considering how many times a broadcast message should be forwarded.

A reliable broadcast (RB) method is proposed in [8], which combines area based and neighbor-based technique of broadcast. Each node gains knowledge of neighbors and maintains neighbor list. The algorithm calculates the relative position of the nodes with respect to broadcast source node. The nodes that are farthest from the source rebroadcasts next. The algorithm tries to minimize the number of rebroadcasts by intermediate nodes and thus reduces message cost.

Reference [9] proposes a method for reduction of broadcast traffic (RBT) in mobile ad hoc networks. It focuses on the fact that communication links in ad hoc networks break frequently due to node mobility. As the nodes move, a node receiving a packet on the boundary of communication range of a transmitter node is allowed to drop the packet, as the receiver may soon move out of the radio range of the transmitter. To approximate the distance between receiver and transmitter, receiver signal strength information is used.

Probabilistic broadcast approaches [10], broadly called gossip, offer a simpler alternative to deterministic approaches. With gossiping, nodes forward packets with a pre-specified probability. The key idea is that when this probability is chosen correctly, the entire network receives the broadcast message with very high probability, even though only a non-deterministic subset of nodes has forwarded the message. Gossiping is a simple solution yet capable of achieving better reliability and load-balancing. However, choosing its correct value is difficult, since it is closely related to network topology information. In absence of topology information, estimating the value of gossip probability is risky. Moreover, the topology of ad hoc networks change from time to time due to link failure and node failure, and therefore a suitably chosen gossip probability may become sub-optimal later. The article in [6], proposes a smart gossip technique which assigns importance to each node in achieving dissemination. The importance of a node increases when other nodes heavily depends on it to disseminate the broadcasted message. The important of a node increases when other nodes heavily depend on it to disseminate a message. The important nodes transmit with a proportionally higher probability. Other nodes that are less crucial for achieving dissemination still transmit for the purpose of reliability but with a lower probability. Initially, when dependencies are not known, gossip probability of each node equals 1. Overtime as nodes learn about their dependencies, the gossip probabilities are refined.

In double covered broadcasting [11], when a sender broadcasts a packet, it selects a subset of 1-hop neighbors as its forward nodes to forward the broadcast based on a greedy approach. The selected forward nodes satisfy two requirements: (1) They cover all the nodes within 2 hops of the sender. (2) The sender's 1-hop neighbors are either forward nodes or non-forward nodes but covered by at least two neighbors, once by the sender itself and once by one of the selected forward nodes. After receiving the broadcast packet, each forward node records the packet, computes its forward nodes and re-broadcasts the packet as a new sender. The retransmissions of the forward nodes are received by the sender as the acknowledgement of receiving the packet. The non-forward 1-hop neighbors of the sender do not acknowledge receipt of the broadcast. The sender waits for a predefined duration to overhear the rebroadcasting from its forward nodes. If the sender fails to detect all its forward nodes retransmitting during this duration, it assumes that a transmission failure has occurred for this broadcast because of the transmission error or because the missed forward nodes are out of its transmission range. The sender then re-sends the packet until all forward nodes are retransmitted or the maximum number of retries is reached. The proposed algorithm utilizes the method that the sender overhears the retransmission of the forward nodes to avoid the ACK implosion problem. Also, the algorithm guarantees that each node is covered by at least two transmissions so that it can avoid a single error due to the transmission collision. Moreover, the algorithm does not suffer the disadvantage of the receiver-initiated approach that needs a much longer delay to detect a missed packet.

## 3. Overview of FLB

Our proposed algorithm FLB works in a clustered environment. For clustering purpose, I have used a multi-hop clustering algorithm based on neighborhood benchmarks (MCNB [12]). This article assumes that all

network links are bidirectional. The score $s_i(t)$ of a mobile node $n_i$, at time t, used to indicate qualification of the node to be a cluster-head, is defined as, $s_i(t)= |D_i(t)| / lf_i$ where $D_i(t)$ is the set of downlink neighbors of $n_i$ at time t and $lf_i$ is the number of link failures encountered by $n_i$ in unit time, indicating link stability of its neighborhood. A node is attached to a cluster provided distance of the new node from head of the cluster is less than or equal to the hop count in the network.

Each cluster has a cluster-head and all cluster members (nodes in a cluster other than the cluster-head) are connected to it. The isolated nodes that are not member of any cluster, are treated as heads of single node cluster. In order to remove redundancy, a constraint is imposed that a node cannot be member of more than one cluster. If source of a broadcast operation is not a cluster-head, it sends the broadcast packet to head of its own cluster. All cluster-heads are connected to each other in single or multi-hop paths. When a cluster-head receives a broadcast packet from its upstream cluster-head, it chooses some gateways or forward nodes to forward the packet to all cluster-heads in its coverage set. The coverage set is updated by excluding the cluster-head sender and those cluster-heads in the senders coverage set that are piggybacked with the broadcast packet. The coverage set of this cluster-head together with its selected forward nodes are piggybacked with the broadcast packet for the forwarding purpose. On the other hand, a cluster-head will do nothing if it receives a duplicate packet. Similarly, cluster members also drop duplicate packets.

A cluster member belonging to the cluster C1, elects its most eligible uplink neighbor among all of its uplink neighbors in cluster C1, by means of recommendation of a fuzzy controller named "Broadcast Neighbor Decider (BND)" which is embedded in every node in the ad hoc network. The parameters of $n_i$ considered by BND of node $n_j$ (here $n_i$ is an uplink neighbor of $n_j$ and both $n_i$ and $n_j$ belong to the same cluster) are residual energy, existing communication load, predictive communication load of $n_i$ and strength of the wireless bond between $n_i$ and $n_j$. The most eligible uplink neighbor of $n_j$ that belongs to the same cluster as $n_j$, is assigned the responsibility of transmitting broadcast message to $n_i$. Design of BND is based on the following heuristics:

i) If a node is already running short of battery power, it should not be assigned the additional responsibility of forwarding broadcast packets to any of its downlink neighbors.

ii) If message queue of a node is almost full and its rate of call arrival is high, then its communication load is huge. As a result, unnecessary delays will be introduced during broadcast operation if nodes like this are elected as most eligible uplink neighbor. The situation will worsen if a) the node has a huge number of uplink neighbors and b) it has already been chosen as most eligible uplink neighbor by a large fraction of its downlink neighbors.

iii) Node $n_i$ will be considered extremely important from the perspective of partition avoidance provided the uplink neighbors of $n_i$ find it difficult to disseminate information to the network without $n_i$ and the downlink neighbors fail to receive information from the network through the nodes other than $n_i$. If the additional responsibility of most eligible uplink neighbor is assigned to such important nodes then their rate of energy depletion will increase resulting to fast exhaustion and network partition, which is not desirable. Hence, the nodes that play important role in maintaining network connectivity are not suitable candidates for being most eligible uplink neighbors of any node. On the other hand, if uplink neighbors of a node $n_j$ has a huge number of downlink neighbors and downlink neighbors of $n_j$ has a huge number of uplink neighbors, then $n_j$ is a good candidate for being most eligible uplink neighbor of some node.

iv) In spite of mobility, the wireless bond between a node $n_j$ and its most eligible uplink neighbor $n_i$ should survive for a significantly long time. Otherwise, $n_j$ will have to frequently elect its most eligible uplink neighbor, increasing complexity of FLB and delay in broadcasting packets.

The observations expressed above are in the form of if-then rules which are the basic unit of fuzzy function approximation. Advantages of fuzzy logic are that it is flexible, conceptually easy to understand and based on natural language. Moreover, it is tolerant of imprecise data and can model non-linear functions of arbitrary complexity. All these encouraged us to design the scheme of FLB using fuzzy logic.

## 4. Parameters of BND

1. The residual energy index $\alpha_i(t)$ of $n_i$ at time t is given by,

$$\alpha_i(t) = (1-e_i(t)/E_i) \quad (1)$$

$e_i(t)$ and $E_i$ specify the consumed energy of $n_i$ till time t and maximum battery power of the same node. From the formulation in (1) it is evident that $\alpha_i(t)$ ranges between 0 and 1. The higher is the value of $\alpha_i(t)$ the more well-equipped is the node to take charge of most eligible uplink neighbor of some node.

2. The uplink neighbor affinity $\beta_{ij}(t)$ of the link from $n_i$ to $n_j$ at time t indicates strength of the wireless bond between those two nodes. If $n_i$ has low velocity relative to $n_j$, then there is high chance that their link will survive for a significantly long time in future. Moreover, the possibility of survival of the link from $n_i$ to $n_j$ increases if $n_i$ has a high radio-range. The situation can be illustrated from figures 1, 2a, 2b, 3a and 3b, based on the assumption that the nodes are moving with uniform velocities.

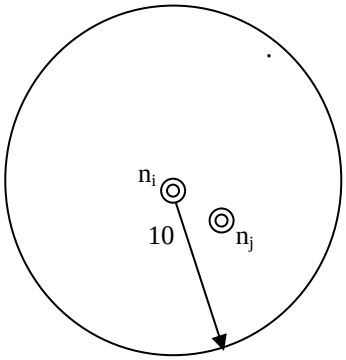



Fig 1

Fig 1: Let the current distance between $n_i$ and $n_j$ be 4 m and radio-range of $n_i$ be 10 m. If the relative velocity of $n_i$ w.r.t. $n_j$ is 2 m/s, then the link between them will survive for (10-4)/2 si.e. 3 s. On the other hand if the velocity of $n_i$ w.r.. $n_j$ be 3 m/s, then the said link will survive for (10-4)/3 s i.e. 2s. Hence low relative velocity of nodes is good for survival of the link between them.

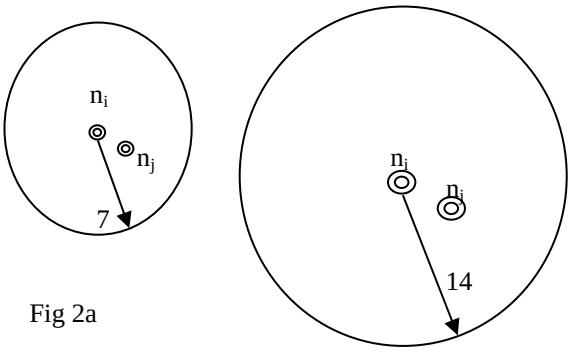



Fig 2a

Fig 2b

Fig 2a and 2b: Let the current distance between $n_i$ and $n_j$ be 5 m and radio-range of $n_i$ be 7 m. If the relative velocity of $n_i$ w.r.t. $n_j$ is 2 m/s, then the link between them will survive for (7-5)/2 s i.e. 1 s. On the other hand if radio-range of $n_i$ is 14 m/s and relative velocity of $n_i$ w.r.t. $n_j$ increases to 3 m/s then the said link will survive for (14-5)/3 s i.e. 3s. Hence high radio-range of a node is good for survival of the link between the node and any of its downlink neighbors.

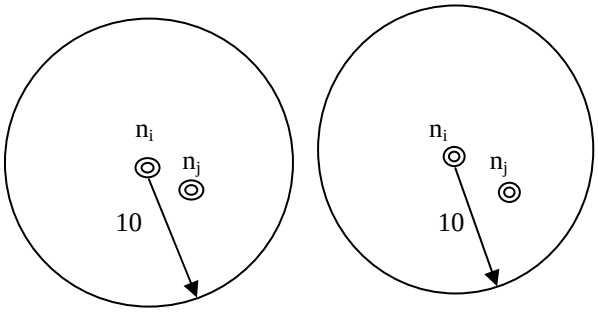



Fig 3a Fig 3b

Fig 3a and 3b: Let the current distance between $n_i$ and $n_j$ be 4 m in fig 3a and 6 m in fig 3b. In both the figures, radio-range of $n_i$ is 10 m. If the relative velocity of $n_i$ w.r.t. $n_j$ is 2 m/s, then the link between them will survive for (10-4)/2 s i.e. 3 s in fig 3a. On the other hand in fig 3b the said link will survive for (10-6)/2 s i.e. 2s. Hence low distance of a node from its downlink neighbor is good for survival of the link between them.

$$\beta_{ij}(t) = (1 - c_{ij}(t)\, r_i)\, (1- d_{ij}(t)/(R_i+1)) \quad (2)$$

Where $c_{ij}(t) = (1\text{-}1/(|v_i(t) - v_j(t)| +1))$
and $r_i = (R_{max}\text{-}R_i+1)/( R_{max}\text{-}R_{min}+1)$

For any node $n_i$, $v_i(t)$ specifies its velocity at time t and $R_i$ specifies its radio-range. Assuming that $R_{min}$ and $R_{max}$ denote the minimum and maximum possible radio-ranges of the network, for any node $n_i$, $R_i$ lies between $R_{min}$ and $R_{max}$. $d_{ij}(t)$ indicates distance of $n_j$ from its uplink neighbor $n_i$ at time t. Magnitude of the relative velocity of $n_i$ w.r.t. $n_j$ or the same of $n_j$ w.r.t. $n_i$, is given by $|v_i(t) - v_j(t)|$.

It may be noted from (2) that $\beta_{ij}(t)$ increases with decrease in $c_{ij}(t)$, $r_i$ and $d_{ij}(t)$. Also $c_{ij}(t)$ decreases with decrease in $|v_i(t) - v_j(t)|$. This rightly models the situation that affinity between a node and its downlink neighbor increases with decrease in their relative velocity. As far as $r_i$ is concerned, it decreases as $R_i$ approaches the upper limit $R_{max}$ of radio-ranges in the network. Hence, $\beta_{ij}(t)$ increases with increase in $R_i$. In the mathematical expression of $r_i$, 1 is added in both numerator and denominator. The reason is that otherwise $r_i$ would have been 0 in the situation $R_i=R_{max}$ and that would nullify the effect of relative velocity of $n_i$ w.r.t. $n_j$ on $\beta_{ij}(t)$, which is not desirable. For any node $n_j$ and its uplink neighbor $n_i$, distance $d_{ij}(t)$ between them at time t must be less than or equal to $R_i$. It is evident from (2) that affinity $\beta_{ij}(t)$ reduces as $d_{ij}(t)$ becomes close to $R_i$ and obtains maximum value if $d_{ij}(t)$ is equal to 0. Please note that 1 is added with $R_i$ in (2) to retail the effects of $c_{ij}(t)$ and $r_i$ on $\beta_{ij}(t)$ when $d_{ij}(t)$ is equal to $R_i$.

Since $c_{ij}(t)$ and $r_i$ are fractions and $d_{ij}(t)$ is less than or equal to $R_i$, so $\beta_{ij}(t)$ ranges between 0 and 1. Values close to 1 emphasize worthiness of $n_i$ as most eligible uplink neighbor of $n_j$.

3. Communication Load $\gamma_i(t)$ of node $n_i$ at time t depends upon the following things:
i) The pending message forwarding load present in message queue of $n_i$ at time t.
ii) The uplink neighbor load of $n_i$ at time t
iii) The number of downlink neighbors that have already chosen $n_i$ as most eligible uplink neighbor
Let AR and N denote the total geographical area of the network and total number of nodes in the network. Then density $\psi$ of nodes in the network is given by,

$$\psi = N /AR \quad (3)$$

Also assume that $U_i(t)$ and $D_i(t)$ denote the set of uplink and downlink neighbors within the same cluster of $n_i$, respectively, of node $n_i$ at time t. Since $R_{max}$ is the maximum possible radio-range of the network, the maximum distance of $n_i$ from any of its uplink neighbors is $R_{max}$. If density of nodes is uniform, then the maximum number of uplink neighbors of any node $n_i$ is $\psi\pi R^2_{max}$. Assume that among $|D_i(t)|$ number of downlink neighbors, $\rho_i(t)$ number of nodes have selected $n_i$ as most eligible uplink neighbor till time t.
The Communication Load $\gamma_i(t)$ is mathematically expressed as,

$$\gamma_i(t) = 1 - [(m_i(t) / M_i ) f_i(t) (\rho_i(t) / |D_i(t)|)]^{1/3} \quad (4)$$

Where $f_i(t) = MIN\{(|U_i(t)| / \psi\pi R^2_{max} ),1\}$
$m_i(t)$ and $M_i$ specify the number of filled locations in message queue of $n_i$ at time t and total number of locations in message queue of the same node. It is quite evident that $\gamma_i(t)$ increases with increase in $m_i(t)$, $|U_i(t)|$ and $\rho_i(t)$ while $m_i(t)$ ranges between 0 and $M_i$ and $\rho_i(t)$ ranges between 0 and $|D_i(t)|$. As far as $|U_i(t)|$ is concerned, it ranges from 0 to N. But it is good for $n_i$ if upper limit of $|U_i(t)|$ is restricted within $\psi\pi R^2_{max}$ which is the maximum under uniform node distribution. If the number of uplink neighbors of $n_i$ increase abruptly, chances of call arrival at $n_i$ in future also increase. MIN is a function that returns the minimum value among its arguments. Please note that if $|U_i(t)|$ is greater than or equal to $\psi\pi R^2_{max}$, then $MIN\{(|U_i(t)|/ \psi\pi R^2_{max}),1\}$ evaluates to 1. Otherwise, $MIN\{(|U_i(t)| / \psi\pi R^2_{max}),1\}$ is a positive fraction. Please note that $|U_i(t)|$ cannot be 0 in a clustered environment, because in a cluster, all members are connected to the cluster. So, there has to be at least 1 uplink neighbor for each cluster member. It is evident from (4) that $\gamma_i(t)$ ranges between 0 and 1. The higher is the value of $\gamma_i(t)$ the more well-equipped is the node to take charge of most eligible uplink neighbor of some node.

4. Connectivity Contribution $\delta_i(t)$ of $n_i$ at time t is formulated as,

$$\delta_i(t) = \delta_{i1}(t) \times \delta_{i2}(t) \quad (5)$$

$$\delta_{i1}(t) = MIN\left[\left\{\prod_{n_j \in U_i(t)} (|D_j(t)| / \psi\pi R^2_j )^{1/|U_i(t)|}\right\}, 1\right] \quad (6)$$

$$\delta_{i2}(t) = MIN\left[\left\{\prod_{n_j \in D_i(t)} (|U_j(t)| /\psi\pi R^2_{max} )^{1/|D_i(t)|}\right\}, 1\right] \quad (7)$$

$\delta_i(t)$ increases with increase in $\delta_{i1}(t)$ and $\delta_{i2}(t)$. $\delta_{i1}(t)$ acquires a high value if the uplink neighbors of $n_i$ at time t are equipped with sufficient number of downlink neighbors at that time and similarly, $\delta_{i2}(t)$ obtains a high value if the downlink neighbors of $n_i$ at time t are equipped with sufficient number of uplink neighbors at that time. For a node $n_j$ with radio-range $R_j$, $\psi\pi R^2_j$ is considered sufficient number for downlink neighbors which is equal to the highest number of downlink neighbors for the radio-circle of radius $R_j$ under uniform node distribution. Similarly, $\psi\pi R^2_{max}$ is considered sufficient number for uplink neighbors which is equal to the highest number of uplink neighbors for any node under uniform node distribution. Please note that for any node $n_j \in U_i(t)$, $D_j(t)$ cannot be empty since it contains at least $n_i$; similarly, for any node $n_j \in D_i(t)$, $U_j(t)$ cannot be empty since it contains at least $n_i$. Values of $\delta_i(t)$ close to 1 increase capacity of $n_i$ as most eligible uplink neighbor of its downlink neighbors.

## 5. Design of Rule Bases of BND

The parameters of BND are divided into crisp ranges and the corresponding fuzzy variables are shown in table 1. Subscripts are omitted for the purpose of simplicity.

Table 1
Crisp Ranges of Parameters and Fuzzy Variables

| Crisp ranges of $\alpha$ | Crisp ranges of $\beta$, $\gamma$, $\delta$ | Fuzzy variable |
|---|---|---|
| 0-0.40 | 0-0.25 | a1 |
| 0.40-0.60 | 0.25-0.50 | a2 |
| 0.60-0.80 | 0.50-0.75 | a3 |
| 0.80-1.00 | 0.75-1.00 | a4 |

According to the study of discharge curve of batteries heavily used in ad hoc networks, at least 40% (fuzzy variable a1 represents the range 0-0.40) of total charge is required to remain in operable condition; 40%-60% (fuzzy variable a2) of the same is satisfactory, 60%-80% (fuzzy

variable a3) is good and the next higher range (i.e. 80%-100% or fuzzy variable a4) is more then sufficient from the perspective of remaining energy. All other parameters follow uniform range distribution between 0 and 1 i.e. (0-0.25 as a1, 0.25-0.50 as a2, 0.50-0.75 as a3 and 0.75-1.00 as a4).Table 2 combines the effects of $\alpha$ and $\beta$ producing temporary output t1. Both are given equal importance since they are equally indispensable for survival of the link from a node to its downlink neighbor. The other parameters contribute to delay-efficiency of the link. The fuzzy composition of t1 and $\gamma$ appears in table 3. In this table, t1 is assigned more importance than $\gamma$ because t1 is a composition of two parameters both of which are more important than $\gamma$. The temporary output t2 generated by table 3 is combined with $\delta$ in table 4 producing final output el of BND.

Table 2
Fuzzy Combination of $\alpha$ and $\beta$ producing output t1

| $\alpha\rightarrow$ $\beta\downarrow$ | a1 | a2 | a3 | a4 |
|---|---|---|---|---|
| a1 | a1 | a1 | a1 | a1 |
| a2 | a1 | a2 | a2 | a2 |
| a3 | a1 | a2 | a3 | a3 |
| a4 | a1 | a2 | a3 | a4 |

Table 3
Fuzzy Combination of t1 and $\gamma$ producing output t2

| t1$\rightarrow$ $\gamma\downarrow$ | a1 | a2 | a3 | a4 |
|---|---|---|---|---|
| a1 | a1 | a2 | a3 | a3 |
| a2 | a1 | a2 | a3 | a3 |
| a3 | a1 | a2 | a3 | a4 |
| a4 | a2 | a3 | a3 | a4 |

Table 4

Fuzzy Combination of t2 and $\delta$ producing output el

| t2$\rightarrow$ $\delta\downarrow$ | a1 | a2 | a3 | a4 |
|---|---|---|---|---|
| a1 | a1 | a2 | a3 | a3 |
| a2 | a1 | a2 | a3 | a3 |
| a3 | a1 | a2 | a3 | a4 |
| a4 | a1 | a2 | a3 | a4 |

If more than one uplink neighbors of a node $n_i$ acquire highest value for el, then any one of those candidates is selected as most eligible uplink neighbor of $n_i$.

## 6. Message Description

FLB requires each node to broadcast HELLO message within its radio-range at regular intervals. If a node $n_j$ exists within the radio-range of another node $n_i$, then $n_i$ is termed as an uplink neighbor of $n_j$. The attributes of HELLO message generated by $n_i$ at time t consists of the following information:

- source node identification number $n_i$
- current timestamp t
- current geographical location $(x_i(t),y_i(t))$ in terms of latitude and longitude
- radio-range $R_i$
- current velocity $v_i(t)$
- the number of downlink neighbors that have already selected $n_i$ as most eligible uplink neighbor
- total number of uplink neighbors in the same cluster
- consumed battery power $e_i(t)$ at time t
- Total battery power $E_i$
- Starting time $t_i$ of operation of $n_i$, in the network

Each node $n_j$ residing within the radio-range of $n_i$, replies with an acknowledgement (ACK) message. Its attributes are as follows:

- source node identification number $n_j$
- destination node identification number $n_i$
- current timestamp t′
- current geographical location $(x_j(t),y_j(t))$ in terms of latitude and longitude
- current velocity $v_j(t)$

Format of a broadcast message initiated by $n_i$ and forwarded by $n_j$ is as follows:

- forwarding node identification number $n_j$
- current timestamp t′
- source identification number $n_i$
- message initiation timestamp $t_{b\_s}$
- current velocity $v_j(t)$

If the link of a node $n_i$ with its most eligible uplink neighbor $n_j$ breaks at time $t_b$, then $n_i$ elects its next most eligible uplink neighbor $n_k$ and sends to $n_k$ a special status message with the following attributes:

- source node identification number $n_i$
- destination node identification number $n_k$
- current timestamp t′
- timestamp $t_b$ of the break

- source identification number and timestamp of initiation of last 3 broadcast message received within timestamp $t_b$ (it is expected that nodes at approximately same distance (in terms of number of hops) from cluster-head receive broadcast messages at approximately same time).

Receiving the status message if $n_k$ finds that some broadcast message was not received by $n_i$, $n_k$ transmits those to $n_i$. Chance of some redundancy exists here if the nodes at approximately same distance (in terms of number of hops) from cluster-head do not receive broadcast messages at approximately same time.

## 7. Algorithm Complexity

- ***HELLO Message Overhead***

In FLB, HELLO messages are transmitted by every node at regular intervals to gather local topology information and elect the most eligible uplink neighbor. Assuming N to be the total number of nodes in the network and Δ to be the average node degree, the HELLO overhead H_OVHD(t) at time t is formulated as,

$$H\_OVHD(t) = N \times \Delta \times (t - t_{start}) / \tau' \quad (8)$$

Where $\tau'$ is the uniform interval between HELLO messages of each node and $t_{start}$ is the starting time of operation of the network.

- ***Redundancy in FLB***

In FLB, a node $n_i$ cannot receive broadcast message from more than one uplink neighbor. Hence, ideally, the redundancy is 0.

- ***100% delivery ratio in FLB***

FLB is based on a clustered architecture and it assumes that all cluster-members are connected to their respective cluster-heads. So, most eligible uplink neighbor $n_k$ of any node $n_i$ must be connected to the cluster-head through some route. $n_i$ will receive the broadcast message as soon as $n_k$ receives it from the cluster-head. So, ideally, the delivery ratio of FLB is 100%.

- ***Complexity of Selecting The Most Eligible Uplink Neighbor***

Assume that the average number of uplink and downlink neighbors of a node be $\Delta'$ and Δ, respectively. The complexities of computing values of input parameters of BND for one uplink neighbor, is O(1). For combining the input parameters, tables 1, 2, 3 and 4 need to be consulted. Table 1 is required for crisp range division and determination of fuzzy variables for those ranges. BND had got 4 input parameters. So, 4 accesses to table 1 is required. Then, during combination of those parameters, exactly one access to each of the tables 2, 3 and 4 is needed. In order to determine el of an uplink neighbor, 7 (i.e. O(1)) table accesses are required. Hence, for $\Delta'$ uplink neighbors, the cost of determining el is $O(\Delta')$. Among all those el's the best one is to be computed. In the best case, el of the uplink neighbor considered first is a4. So, the best case cost is 1. On the other hand, the corresponding worst case cost is $(\Delta'-1)$ (i.e. $O(\Delta')$) , where el of first $(\Delta'-1)$ uplink neighbors is not a4. So, the overall complexity of selecting the most eligible uplink neighbor is $O(\Delta')$.

- ***Cost of intra-cluster and inter-cluster communication***

Cost of inter-cluster communication increase if the number of clusters increase or the size of clusters decrease. Decrease in the size of clusters will reduce the cost of intra-cluster communication. Let, the total number of clusters in the network be denoted as cls_num. Also assume that hlim and clim denote the maximum distance of a cluster member from its cluster-head in terms of number of hops and maximum number of nodes in a cluster, respectively. Then,

$$cls\_num \times clim = N \quad (8)$$

i.e. $cls\_num = N / clim$

Cost of inter-cluster communication is given by O(cls_num). Cost of intra-cluster broadcast and unicast communication are O(clim) and O(hlim) respectively. hlim is less than or equal to the hop count H of the network. If clim is set to $\sqrt{N}$, then cost of both inter-cluster communication and intra-cluster broadcast becomes $O(\sqrt{N})$. On the other hand, if clim is set to $N^{1/3}$, then cost of inter-cluster communication and intra-cluster broadcast are $O(N^{2/3})$ and $O(N^{1/3})$, respectively. So, clim is the handle that is used to obtain a trade-off between the costs of inter-cluster and intra-cluster communication.

## 8. Simulation Results

I evaluate the performance of FLB, using the network simulator ns-2. Except FLB, I implement the protocols tree-based broadcasting (TBB), reduction of broadcast traffic (RBT) approach, reliable broadcasting (RB) and density-based flooding (DBF). The ns-2 is a discrete event simulator developed by the University of California at Berkeley and VINT project [12]. For the purpose of studying multi-hop ad hoc networks, it has been modified and extended with mobile wireless modules by the CMU Monarch project [12]. This simulator has been used to evaluate the performance of ad hoc routing protocols. Each mobile node has a position and velocity. In different

simulation runs, nodes move according to the "random waypoint", "random walk" and "gauss-markov" model. In random waypoint model, each node begins operation by remaining stationary for PAUSE_TIME seconds (its value is mentioned in table 1). It then selects a random position in the space and moves to that position at a speed distributed uniformly between 0 and MAX_SPEED. When it reaches the destination, a new round of pause/ move is repeated. The random walk model was originally used to emulate the unpredictable movements of particles in physics, also referred to as Brownian motion. Random walk model is very similar to random waypoint mobility model because the node movements have strong randomness in both models. The random walk model may be thought of as a specific kind of random waypoint model with PAUSE_TIME 0 seconds. On the other hand, in gauss-markov mobility mode, the velocity of a node is assumed to be correlated over time and modeled as a gauss-markov stochastic process. The parameters of simulation are shown in table 5.

The performance metrics I observe are:-

- Broadcast cost – It is the normalized average cost to deliver a broadcast message to all nodes in the network. It is defined as TOT_MSG / (TOT_BRC_SRC * N) where TOT_MSG is the total number of messages transmitted by all nodes in the network, including the control messages, TOT_BRC_SRC is the total number of user messages generated by the broadcast sources and N is the total number of mobile nodes. TOT_MSG is also a metric of the bandwidth consumed in broadcast.
- Delivery ratio – It is defined as TOT_RECV / (TOT_BRC_SRC * (N -1 )) where TOT_RECV is the total number of non-duplicate messages received by users. The delivery ratio reveals the robustness of the simulated protocol. In the ideal case, the delivery ratio will be 1.
- Delay - It is defined as (for all broadcast messages $\sum$ (MSG_END_TIME - MSG_START_TIME) )/ (TOT_MSG) where TOT_MSG is the total number of broadcast messages transmitted. MSG_END_TIME is the timestamp when a broadcast operation completed i.e. reached to all nodes in the network. Similarly, MSG_START_TIME is the timestamp when a broadcast message was transmitted by its source i.e. when the broadcast operation initiated.

The performance of FLB is compared with the performance of some state-of-the-art broadcast protocols, namely RB, RBT, TBB and DBF. The corresponding graphical representations appear in figures 1 to 6.

I have already discussed the fact that in FLB, each node selects the most efficient uplink neighbor considering residual energy, strength of wireless bond between a node and its most efficient uplink neighbor, communication load and contribution in maintaining connectivity in the network. Only the selected uplink neighbors rebroadcast a message while the others drop the packet after receiving it. Hence, redundancy in FLB is 0 and the broadcast cost is much lesser compared to the other protocols mentioned above. TBB does not suffer from much redundancy because it is based on tree-structured nodes, but as the number of nodes increase, the phenomenon of link breakage becomes more frequent because more links are there to be maintained. As a consequence, the broadcast tree structure requires modification increasing the broadcast cost. On the other hand, links between a node and its most efficient uplink neighbor in FLB are stable, reducing the overhead of frequent re-election of most eligible uplink neighbor. The improvement can be noticed from figures 1 and 2 where broadcast cost is measured with respect to total number of nodes and number of broadcast source, respectively. It is quite evident that for all the above-mentioned protocols, broadcast cost increases with increase in number of nodes and broadcast sources, with the reason being increased signal collision in the network. But the dependence of broadcast cost on node mobility should be discussed separately. TBB creates a tree for broadcasting whose links break frequently if node mobility increases. This requires restructuring of the tree by exchanging some more messages. Hence, broadcast cost for TBB increases with average node velocity. But for others, the cost is steady. This is shown in figure 3.

Since, FLB is power aware, energy depletion in nodes is quite balanced. As a result, the chances of network partitioning get reduced. Also mobility awareness brings stability in relationship between a node and its most efficient uplink neighbor. All these contribute to produce packet delivery ratio as high as 99.98%. The delivery ratio increase for all the protocols, when the network scales large. This is because the network becomes dense with the increase of node number in a fixed size area and a mobile node is more likely to be covered by a broadcast relay gateway. In figure 5, delivery ratio is measured with respect to number of sources. As the number of broadcast sources increase, a huge number of messages need to be forwarded network wide. This, in turn, generates signal contention and collision resulting in the drastic drop in delivery ratio. Since FLB is power and mobility aware and does not suffer from redundancy, it can efficiently resist the drop for a longer time duration than RB, RBT, TBB and DBF. The phenomenon is illustrated in figure 5. Figure 6 illustrates the dependence of delivery ratio on

average node velocity. Since broadcast cost increases in TBB with increase in node velocity, delivery ratio decreases with node velocity due to signal collision and high rate of energy depletion of nodes. For the other protocols the delivery ratio remains steady with node mobility.

Table 5
Simulation Parameters

| Parameter | Value |
|---|---|
| Network Area | 500 × 500 $m^2$ in first ten runs, 1000 × 500 $m^2$ in next ten runs, 1000 × 1000 $m^2$ in last ten runs |
| Transmission Range | 10 – 50 m in first ten runs, 30 – 100 m in next ten runs, 10 – 100 m in last ten runs |
| Interval between consecutive HELLO messages | 20 seconds for first ten simulation runs, 30 seconds for next ten and 45 seconds for last ten simulation runs |
| Number of nodes | 30 - 300 |
| MAC layer | IEEE 802.11g |
| PAUSE_TIME | 20 seconds |
| Traffic type | Constant bit rate (128 kbps/second) |
| Maximum number of retries before an acknowledgement is obtained | 4 |
| Packet Size | 64 bytes in first ten runs, 128 bytes in next ten runs, 256 bytes in last ten runs (in different simulation runs) |
| Bandwidth | 1- 4 Mbps in first ten runs, 2 – 7 Mbps in first ten runs, 1-10 Mbps in last ten runs |
| Mobility model | Random waypoint mobility model in first 10 runs, Random walk mobility model in subsequent 10 runs and Gaussian model in last 10 runs |
| Simulation Time | 1000 seconds for each run |

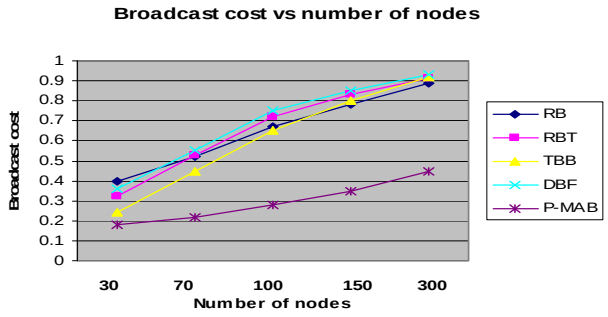



Figure 1: Graphical demonstration of broadcast cost vs number of nodes

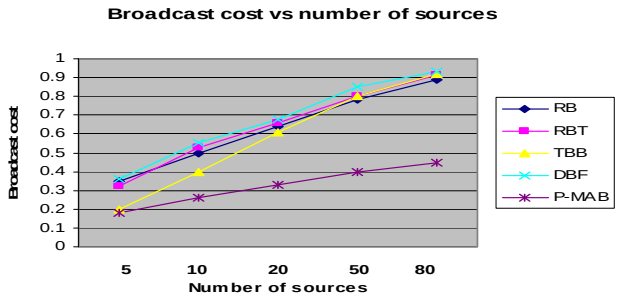



Figure 2: Graphical demonstration of broadcast cost vs number of sources

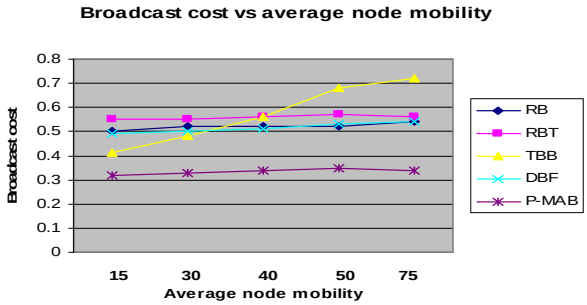



Figure 3: Graphical demonstration of broadcast cost vs average node velocity in meter/second

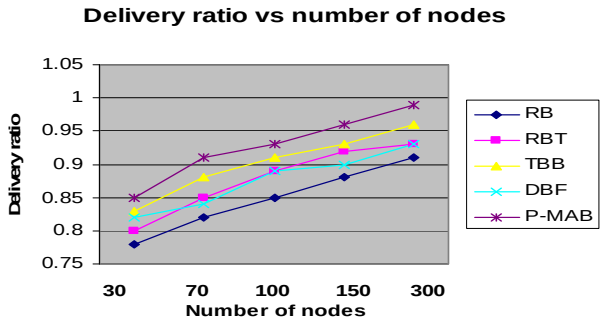



Figure 4: Graphical representation of delivery ratio vs number of nodes

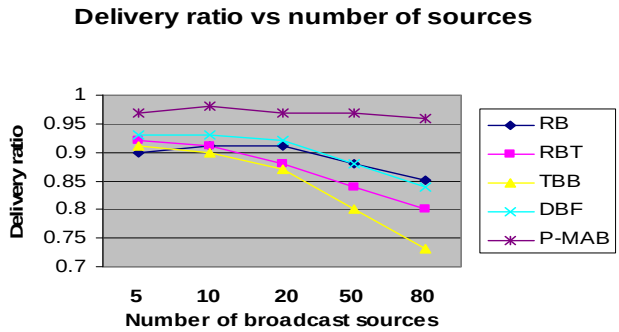



Figure 5: Graphical representation of delivery ratio vs number of source

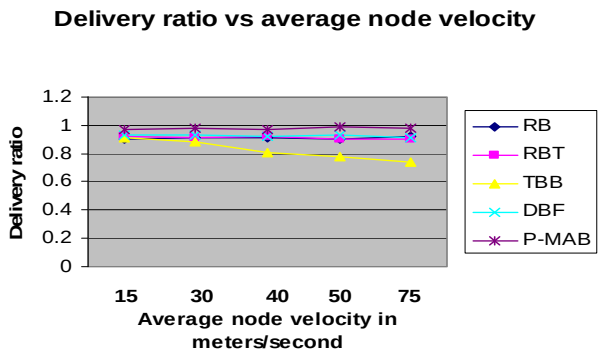



Figure 6: Graphical representation of delivery ratio vs average node velocity

## 9. Conclusion

This paper presents a new approach for efficient broadcasting in mobile ad hoc networks. The proposed protocol called FLB is both power and mobility aware. Each node selects its most efficient uplink neighbor and receives broadcast message from only that neighbor. It minimizes the broadcast redundancy and also saves the network bandwidth. Most efficient uplink neighbor is elected by considering residual energy and link stability. This brings power and mobility awareness in the protocol.

## References

[1] Anuradha Banerjee, Paramartha Dutta, "A Survey of Unicast Routing Protocols in Mobile Ad Hoc Networks", International Journal of Advances in Science and Technology, vol. 2, no. 10, 2010

[2] D David B. Johnson. "Routing in Ad Hoc Networks of Mobile Hosts". Proceedings of the Workshop on Mobile Computing Systems and Applications, pp. 158–163, IEEE Computer Society, Santa Cruz, CA, December 1994

[3] Ian D. Chakeres et. Al, "AODV routing protocol implementation design", In Proceedings of 2nd International Workshop on Wireless Ad Hoc Networking **...** Notification (WWAN), March 2005

[4] Lee Sung-Ju, Su W., M. Gerla, "On-demand Multicast Routing Protocol (ODMRP) for Ad Hoc Networks", June 1999

[5] Ho. C, Obraczka K, Tsudil G., "Flooding for Reliable Multicast in Multi-hop Ad Hoc Networks", In Proceedings of 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIAL M'99), August 1999

[6] N. Karthikeyan, V. Palanisamy, K. Duraiswamy, "Optimum Density Based Model for probabilistic Flooding Protocol in Mobile Ad Hoc Network", European Journal of Scientific research, vol. 39 no.4, pp. – 577-588, 2010

[7] L. Tan, X. Zhan, J. Lie, F. Zhao, "A Novel Tree-based Broadcast Algorithm for Wireless Ad Hoc Networks", International Journl of Wireless and Mobile Computing, vol. 1 no. 2, pp. 156 – 162, 2006

[8] S.V.M.G. Bavithiraja, R. Radhakrishnan, "A New Reliable Broadcasting Method in Mobile Ad Hoc Networks",International Journal of Computer Science and Network Security, vol. 9, no. 4, pp. 340-349, 2009

[9] A. Shukla, "On the Reduction of Broadcast Traffic in Mobile Ad Hoc Networks", In Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, September 2007

[10] G. wang, D. Lu, W. Jia and J. Chao, "Reliable gossip based broadcast protocol in mobile ad hoc networks", Lecture Notes in Computer Science 2005, Volume 3794/2005, 207-218, DOI: 10.1007/11599463_21

[11] W. Lou, J. Wu, "Double Covered Broadcast (DCB): A Simple Reliable Broadcast Algorithm in Ad Hoc Networks", http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.2687, 2004

[12] http://isi.edu/nsnam/ns

Anuradha Banerjee is presently working as assistant professor at Kalyani Govt. Engg College, India. She obtained her B.E. in Computer Sc. and Technology from Bengal Engineering College, Sibpur in 2001 and pursuing Ph.D. in ad hoc networks. She has 17 international journal publications to her credit and 9 conference publications. Her areas of interest are ad hoc networks, neural networks and artificial intelligence.

# Employee Likelihood of Purchasing Health Insurance using Fuzzy Inference System

**Lazim Abdullah[1] and Mohd Nordin Abd Rahman[2]**

**[1] Department of Mathematics, University Malaysia Terengganu, Kuala Terengganu, 21030, Malaysia**

**[2] Faculty of Informatics, University of Sultan Zainal Abidin, Kuala Terengganu, 21030, Malaysia**

**Abstract**

Many believe that employees' health and economic factors plays an important role in their likelihood to purchase health insurance. However decision to purchase health insurance is not trivial matters as many risk factors that influence decision. This paper presents a decision model using fuzzy inference system to identify the likelihoods of purchasing health insurance based on the selected risk factors. To build the likelihoods, data from one hundred and twenty eight employees at five organizations under the purview of Kota Star Municipality Malaysia were collected to provide input data. Three risk factors were considered as the input of the system including age, salary and risk of having illness. The likelihoods of purchasing health insurance was the output of the system and defined in three linguistic terms of 'Low', 'Medium' and 'High'. Input and output data were governed by the Mamdani inference rules of the system to decide the best linguistic term. The linguistic terms that describe the likelihoods of purchasing health insurance were identified by the system based on the three risk factors. It is found that twenty seven employees were likely to purchase health insurance at 'Low' level and fifty six employees show their likelihoods at 'High' level. The usage of fuzzy inference system would offer possible justifications to set a new approach in identifying prospective health insurance purchasers.

***Keywords:*** *Health Insurance, Fuzzy rule, Risk factor, Fuzzy logic, Likelihoods*

## 1. Introduction

The area of insurance was introduced since three decades ago and now very often becomes part of individual's life. Insurance plays very important industry as it captures the future unpredictable events. It is defined as a form of risk management primarily used to hedge against the risk of contingent loss. There are many type of insurance and one of the most common insurances in health concern societies is health insurance. The term health insurance is generally used to describe a form of insurance that pays for medical expenses. In a modern era today, peoples are more concern about their health care and protection of their wealth. This situation makes health insurance has become one of a must buy insurance coverage. It was reported in Australia that between 1997 and 2001, the health insurance market underwent a rapid transformation as these reforms have increased the insurance coverage amongst order people [1]. Healthcare insurance enables access to care by protecting individuals and families against the high and often unexpected costs of medical care. The 2002 Population Survey reports that nearly eighty three percent of the under-age-sixty-five population in the United States had health insurance. More than three-quarters of these people had coverage through an employer, fewer than 10 percent purchased coverage on their own, and the remainder had coverage through a government program [2].

As to meet health care of the present markets, insurance industries have identified several key factors that may affect insurance prices. A recent study by PricewaterhouseCoopers [3] examining the drivers of rising health care costs in the United States. The report pointed to increased utilization of health care costs was created by increased consumer demand, new treatments, and more intensive diagnostic testing, as the most significant. These significant demands are definitely relate to factors that influence people in purchasing health insurance. The risk factors of age, lifestyle, health conditions and purchasing power, among others, are all affecting to the demand for health insurance. Risk of illness and the attendant cost of care lead to the demand for health insurance. Besides consumers' health and wealth conditions, claim provisions and attractive benefits to claimers are also important to motivate people in purchasing health insurance. However, the affects of

these factors toward likelihoods of purchasing insurance are still not conclusive. This view is supported by McLaughlin, et al. [4] who writes that although obtaining health insurance is voluntary in the U.S., surprisingly little is known about the factors that determine individual to obtain health insurance.

Many approaches have been proposed to understand performance of purchasing health insurance. In Taiwan, for example, performance of health insurance industry, was investigated by Liu and Chen [5] using a survey and expenditure data. They investigate the factors influencing the probability and amount of private health insurance purchased using a two-part statistical model of logistic and ordinary least squares regressions. However, unpredictable future and uncertain behaviors of insurance do not make it advisable to forever rely on statistical approaches. Alternatively, approaches using fuzzy sets theory have been flourished in insurance analysis as an alternative or complement to statistical approaches. Ostaszewski [6] was among the first to suggest the use of the *c*-means algorithm for classification in an insurance context. Cummins and Derrig [7] used fuzzy approach to forecasts of automobile bodily injury liability pure premiums. Besides forecasting for premium using fuzzy multiple criteria, fuzzy sets have been used in classifications. Horgby [8] describes how to classify risks by using a fuzzy inference methodology. By defining risk factors as fuzzy sets, it is shown that an insurer can utilize multiple prognostic factors that are imprecise and vague. Apart from forecasting and classification, fuzzy set theory has also been used in evaluation for purchasing insurance. Chin et al, [9] proposed an evaluation model for purchasing life insurance and annuity insurance using a combination of pair-wise comparisons of analytical hierarchy process and fuzzy logic. Four factors were considered as the inputs of the proposed model including age, annual income, educational level and risk preference. The significant role of fuzzy logic in insurance was stressed by Shapiro [10]. He asserts that insurance industry has numerous areas with potential applications for fuzzy logic. These include classification, underwriting, projected liabilities, fuzzy future and present values, pricing, asset allocations and cash flows, and investment. Given these potentials and the impetus on fuzzy logic during the last decade, it is not surprising that a number of fuzzy logic studies have focused on insurance applications. Fuzzy logic is seemed as a tool in insurance pricing decisions that consistently consider vague data. However, far too little attention has been paid to classify the impact of risk factors to likelihood of health insurance purchasing. As an initiative to empower the decision tool using fuzzy logic, this paper proposes another study of fuzzy logic to health insurance purchasing potential. The whole package operations of fuzzy logic have been translated magnificently into fuzzy inference system (FIS). FIS based on fuzzy rules has been applied to numerous engineering applications such as control, signal processing, and pattern classification problems [11][12].

The notion of fuzzy sets in forecasting, classifications and evaluations of insurance fraternity was further explored. The possibility of extending fuzzy inference to classify likelihoods of purchasing health insurance is always plausible. Likelihood of purchasing health insurance based on their risk factors is intended to investigate in this paper. Specifically, this paper aims to propose the likelihoods of purchasing health insurance in form of linguistic terms using FIS. This paper is organized as follows. As to make this paper self-contained, Section 2 describes a brief introduction of FIS. A case study of likelihoods of purchasing health insurance using FIS among selected employees is explained in Section 3. Finally this paper ends with conclusions in Section 4.

## 2. A Brief of Fuzzy Inference System

Fuzzy inference system is a popular methodology for implementing fuzzy logic. FIS is one of the most famous applications of fuzzy logic and fuzzy sets theory [13]. It is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. FIS is sometimes called fuzzy reasoning or approximate reasoning. It is used in a fuzzy rule to determine the rule outcome from the given rule input information. Fuzzy rules represent control strategy or modeling knowledge/experience. When specific information is assigned to input variables in the rule antecedent, fuzzy is needed to calculate the outcome for output variables in the rule consequence.

FIS are also known as fuzzy rule-based systems, fuzzy expert systems, fuzzy models, fuzzy associative memories, or fuzzy logic controllers when used as controllers. The main components of the system include fuzzification interface, inference engine and deffuzification. Basic structure of FIS that comprises three components and rules can be seen in Fig 1.

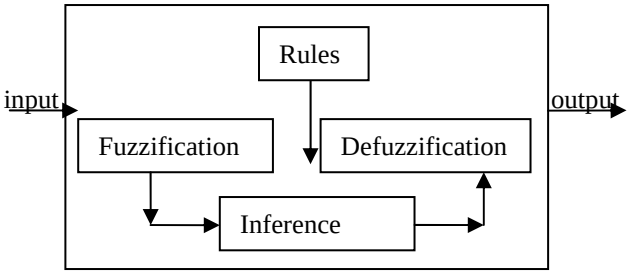


Fig 1 Basic structure of a fuzzy inference system

FIS can be envisioned as involving a knowledge base and a processing stage. The knowledge base provides membership functions and fuzzy rules needed for the process. In the processing stage, numerical crisp variables are the input of the system. These variables are passed through a fuzzification stage where they are transformed to linguistic variables, which become the fuzzy input for the inference engine. This fuzzy input is transformed by the rules of the inference engine to fuzzy output. These linguistic results are then changed by a defuzzification stage into numerical values that become the output of the system. Therefore creating decision using FIS may involves several steps. The steps in FIS are used to test a case study of likelihoods in purchasing health insurance. The steps specifically tailored to the objective of this paper are explained in Section 3.

## 3. A Case of Health Insurance Purchasing

The FIS is tested to one hundred and twenty eight employees who are working at government and public sectors in Kota Star Municipality of Peninsular Malaysia. Three important input variables that may influence the likelihoods of purchasing health insurance are identified. In this case study, the input variables are age, salary and risk of illness. The system uses Mamdani inference [1] which allows a system to take in a set of crisp input values and apply a set of fuzzy rules to those values, in order to derive a single, crisp, output value. The following steps are executed to obtain likelihoods in purchasing health insurance.

*Step 1: Defining input and output*

The factors of likelihoods of purchasing insurance are become the input of this Mamdani inference and the output of the system is likelihoods of purchasing health insurance. Fig 2 illustrates how the inputs related to the health insurance are being processed to create the output.

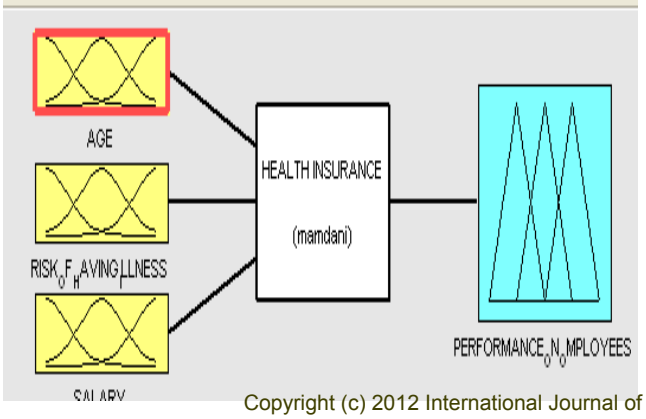



Fig 2 Input and output of the system

Likelihood of purchasing health insurance is labeled as 'performance on employees' in the system as to indicate the employees' capability in purchasing health insurance. Based on the defined system functional and operational characteristics, input crisp data from this experiment are needed to fuzzify.

*Step 2: Defining Fuzzy Sets for System Variables*

System variables need to fuzzify in order to obtain fuzzy membership. The system recognizes the input and output variables and defines its memberships. Memberships for risks of illness, for example are defined in three linguistic terms, 'High', 'Medium' and 'Low'.

*Step 3: Defining Fuzzy Rules*

The next step is defining the If-Then rules to describe system behavior. The rules are designed as to describe the importance of the factors on employees over the possibility of purchasing health insurance. At this step, data from employees are entered into the system. Based on the expert knowledge, this study expresses the problem in terms of logical rules.

For examples, if three respondents were considered, then the rules are given as follows.

Rule 1 (respondent 1) IF age is 25 years old AND the risk of having illness is low AND his salary is RM 2678.00 THEN his likelihoods to buy health insurance is medium.

Rule 1 (respondent 2) IF age is 47 years old AND the risk of having illness is high AND his salary is RM 3472.00 THEN his likelihoods to buy health insurance is high

Rule 1(respondent 3) IF age is 31 years old AND the risk of having illness is medium AND his salary is RM 720.00 THEN his likelihoods to buy health insurance is low

These fuzzy rules are executed for all respondents. Part of the fuzzy rules is shown in Fig 3.

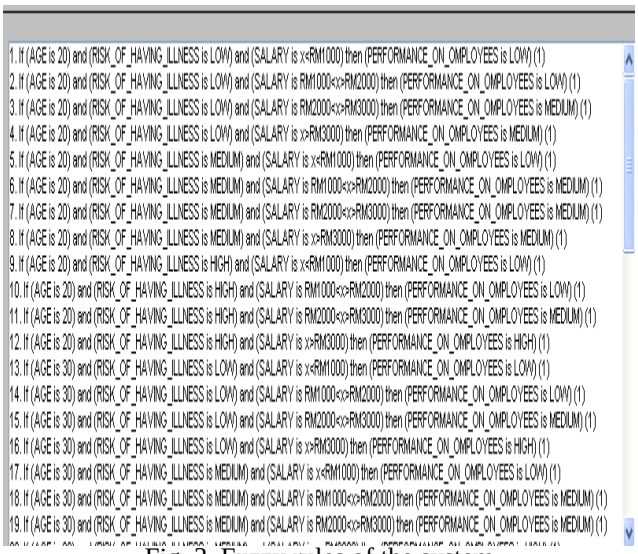



Fig. 3 Fuzzy rules of the system

The inference rules set the premise to create output. The output, then need to defuzzify in order to obtain crisp value.

*Step 4: Defuzzification*

Finally defuzification step is needed to convert all input data into three linguistic terms that can be used to observe the likelihoods of purchasing insurance. The defuzzification process transforms the fuzzy set into a crisp value that is meaningful to end-user. For example, if respondent's age is 43 years and his risk of having illness is medium and his monthly salary is RM 2800.00 then the defuzzification result shows the output is 52. Thus based on the defined output, the likelihood of the employee over the capability of purchasing health insurance is 'medium'. Part of the processes is shown in Fig 4

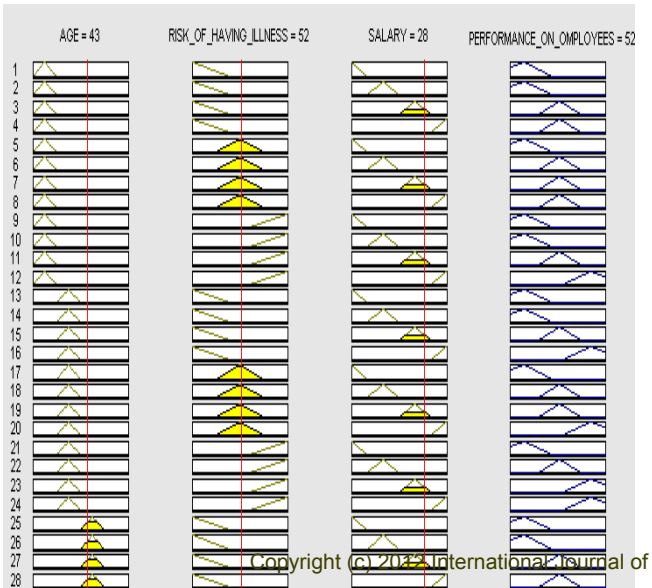



Fig 4 Defuzzification Process

Results for the rest of employees can be obtained with the similar fashion. The likelihoods of purchasing health insurance and the risk factors are partly shown in Table 1.

Table 1 Likelihoods and risk factors of purchasing health insurance

| Employee | Age (year) | Risk of illness | Salary (RM) | Likelihoods |
|---|---|---|---|---|
| 1 | 43 | Medium | 2800 | Medium |
| 2 | 56 | Medium | 3500 | High |
| 3 | 38 | High | 3060 | High |
| 4 | 22 | Low | 870 | Low |
| 5 | 28 | Low | 1210 | Low |
| 6 | 23 | Low | 1610 | Medium |
| 7 | 24 | Low | 1864 | Medium |
| 8 | 37 | Low | 1280 | Low |
| … | … | … | … | … |
| … | … | … | … | … |
| … | … | … | … | … |
| 128 | 28 | Low | 2265 | Medium |

The system ultimately decides the likelihood of purchasing health insurance in one out of three linguistic terms for each employee. The likelihoods of purchasing health insurance are either 'Low', 'Medium' or 'High' . The decisions obtained are based on the three input variables and one output variable governed by the fuzzy rules in FIS. Summarily the likelihoods of purchasing health insurance for one hundred and twenty eight employees are tabulated in Table 2.

Table 2 Number of employee and their likelihood

| Likelihoods | Number of employee | Percentage |
|---|---|---|
| High | 56 | 43.8 |
| Medium | 45 | 35.1 |
| Low | 27 | 21.1 |

The descriptive percentage analysis shows that the likelihood of 27 employees (21.1%) to purchase health insurance is 'Low' and the likelihood of 45 employees (35.1%) is 'Medium'. The likelihood of 56 employees (43.8%) is 'High'. Based on the three inputs, this study has shown that FIS successfully classified the likelihoods of employees into the three linguistic of 'High', 'Medium' and 'Low'. The results of this study indicate that it is very difficult to determine the risk factors that can predict the likelihood of purchasing health insurance.

## 4. Conclusions

An important element in determining the likelihoods of purchasing health insurance is a method which can take into account the multi factors. The method should establish a decision to reflect the contribution of each accounted factor. Furthermore the method should be practical, direct analysis and the most important is the results are easily understandable. In this paper, the fuzzy rules based method to classify the likelihoods of purchasing health insurance was utilized. The system uses fuzzy inferences that can encode the researchers' expertise to reach decision. Three risk factors that affect the potential of purchasing health insurance were considered including age, salary and risk of illness. The system recognized the likelihoods in one out of three linguistic terms intelligently adjusted by the system. The results reveal the effectiveness of the system in identifying the likelihoods of purchasing health insurance. The system can be used as a tool by insurance company in the process of indentifying prospective purchasers. Future research can be extended to rank the risk factors in accordance with employees' preferences using other intelligent methods such as analytic hierarchy process or any other non parametric tests.

## References

[1] Temple J B . Health insurance reform and older Australians. *Australasian Journal on Ageing*, Vol. 25 , 2006, pp.63–68.

[2] Fronstin, P. Sources of Health Insurance and Characteristics of the Uninsured: Analysis of the March 2003 Current Population Survey, EBRI Issue Brief, no. 264 (Washington, D.C.: Employee Benefit Research Institute, 2003).

[3]PricewaterhouseCoopers The Factors Fueling Rising Healthcare Costs for America's Health Insurance Plans, 2006, accessed 2007-10-08.

[4] McLaughlin, C.G, S.E. Crow, M Harrinington,.and H. Kuttner. Causes and Consequences of Lack of Insurance: Gaps in our Knowledge. In Health Policy and the Uninsured, edited by C. G. McLaughlin, pp. xiii-xxv. Washington D.C.:The Urban Institute Press. 2004.

[5] Liu, T.C. and C.S.Chen, An analysis of private health insurance purchasing decisions with national health insurance in Taiwan, *Social Science and Medicine*, Vol. 55, 2002, pp.755-774.

[6] Ostaszewski, K. *Fuzzy Set Methods in Actuarial Science.* Schaumburg, IL Society of Actuaries. 1993.

[7] Cummins, J D. and R.A Derrig. Fuzzy Trends in Property-Liability Insurance Claim Costs, *Journal of Risk and Insurance*, Vol. 60**,** 1993, pp. 429-465.

[8] Horgby, P.J. Risk Classification by Fuzzy Inference. *Journal of The Geneva Papers on Risk and Insurance, Theory*, Vol 23 , 1998, pp. 63-82.

[9] Chin, S. H., J.L.Yu, and C. L. Che, An evaluation model for determining insurance policy using AHP and fuzzy logic: case studies of life and annuity insurances, *Proceedings of WSEAS International Conference on Fuzzy Systems*. Vol.8, 2007, pp. 126-131.

[10]Shapiro A.F. Fuzzy Logic in Insurance. *Journal of Insurance: Mathematics and Economics*, Vol. 35, 2004, pp.399–424.

[11]M. Setnes and H. Roubos, Ga-fuzzy modeling and classification: Complexity and performance, IEEE *Transactions on Fuzzy Systems*, Vol. 8, No.5, 2000, pp. 509–522.

[12] H. Wu and J.M. Mendel, Binary classification of ground vehicles based on the acoustic data using fuzzy logic rule based classifiers, Tech. Rep. 356, USC-SIPI, 2000.

[13]L. A. Zadeh, Fuzzy sets, *Information Control*, Vol. 8, No.3, 1965, pp. 338–353.

[14]E. H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Machine Studies*, Vol. 7, 1975, pp. 1–13.

**First Author** Lazim Abdullah is an Associate Professor at the Department of Mathematics, Faculty of Science and Technology, University Malaysia Terengganu. He holds a B.Sc (Hons) in Mathematics from the University of Malaya, Kuala Lumpur in June 1984 and the M.Ed in Mathematics Education from University Sains Malaysia, Penang in 1999. He received his Ph.D. from the University Malaysia Terengganu, (Information Technology Development) in 2004. Currently, he is Head of E-learning Unit at the University Malaysia Terengganu. His research focuses on the mathematical theory of fuzzy sets and its applications in social ecology, environmental sciences, health sciences, and education. He is interested in the measurement of social indicators, an index of health and educational constructs using computational intelligence approaches. Analyzing data with computing software is an advantage to him because of his direct involvement in undergraduate teaching in the subject of biostatistics and data analysis. Currently he is a member of editorial boards to several international journals related with computing and information technology. Besides, he has been reviewed articles of a number of local and international journals, member of scientific committees of several symposia and conferences at national and international levels. He is an associate member IEEE Computational Intelligence Society, a member of Malaysian Mathematical Society and a member of the Institute of Advanced Scientific Research.

**Second Author** is an Associate Professor at Faculty of Informatics, University of Sultan Zainal Abidin. He is a member of the IEEE and the IEEE Computer Society.

# Robust Iris Recognition Based on Statistical Properties of Walsh Hadamard Transform Domain

**Sunita V. Dhavale**[1]

[1] **Department of computer Engineering, Defence Institute of Advanced Technology**
**Girinagar, Pune-411025, Maharashtra state, INDIA.**

**Abstract**

In this paper, a new approach of iris image feature extraction technique based on the statistical properties of Walsh Hadamard Transform (WHT) domain is proposed. A Canny Edge Detection followed by Hough Transform is used to detect the iris boundaries in the digital image of an eye. The segmented and normalized iris region is divided into 8x8 non-overlapping blocks and WHT is applied to each block. Unique iris features are obtained by computing mean value of energy (MVE) and mean value of standard deviations (MSD) of WHT coefficients. The energy-compaction characteristics of WHT are used to capture iris texture variations. Fast Walsh Hadamard Transform Algorithm is used to reduce the computational time. The features extracted by the WHT domain are used to generate unique encoded binary image and corresponding unique binary bit stream/code is constructed. In order to reduce the size of the database, this binary bit stream instead of binary image is stored in database for matching purpose. Further to increase the security of the system, the bit stream obtained is first encrypted using the user key obtained from user password and then the encrypted bit pattern template is stored. Experimental results on Bath University Iris Database reveal that the proposed iris matching scheme provides results comparable to those of recent methods and is also computationally effective.

***Keywords:*** *Iris recognition, Walsh Hadamard Transform, biometrics, human identification, image preprocessing.*

## 1. Introduction

Automatic reliable personnel identification systems using biometrics have received a great importance in the past few years. Biometric refers to a science of analyzing human physiological or behavioral characteristics for security purposes. The Biometric characteristics cannot be faked, forged, guessed and stolen easily. One need not remember his/her biometric traits [1].

Iris is the round contractile membrane of the eye suspended between cornea and lens which is perforated by the pupil. Iris begins to form during gestation and by the eighth month of the pregnancy it gets completely formed. The iris of the human eye is so unique that no two irises are alike, even among identical twins or even between the left and right eye of the same person, in the entire human population. Also changing iris pattern of any person without surgery with high risk is impossible. Thus it is considered as one of the most reliable biometric in case of biometrics-based identification/recognition systems [2, 6].

A typical iris recognition system involves four main modules. The first module, image acquisition deals with capturing sequence of iris images from the subject using cameras and sensors. The second module, preprocessing involves various steps such as iris liveness detection, pupil and iris boundary detection, eyelid detection and removal and normalization. Several methods like Hough transformation, integro-differential operator, gradient based edge detection are used to localize the portions of iris and the pupil from the eye image. It is essential to map the extracted iris region to a normalized form. The iris localization methods are based on spring force, morphological operators, gradient, probability and moments. The third module, feature extraction identifies the most prominent features for classification. The features are encoded to a format suitable for recognition. The fourth module, recognition achieves result by comparison of features with stored patterns [6, 11].

A major approach for iris recognition today is to generate feature vectors corresponding to individual iris images and to perform iris matching based on different metrics [15]. One of the difficult problems in feature based iris recognition is that, the speed of matching is significantly influenced by time required for feature extraction process, size of the template database stored, format of the template database etc. Thus fast, robust and secured implementation techniques are needed.

The rest of this paper is organized as follows. Section 2 provides overview of related works. Section 3 provides the outline of the proposed algorithm for Iris Recognition system. Experimental results are compared with the results of previous works in Section 4; followed by the conclusions in Section 5.

## 2. Related Works

Daugman's system is the first known algorithm for iris recognition [1]. It consists following major stages:

1. Pre-processing stage: To detect the edges of the pupil and iris and to locate the position of the iris within the image with an integro-differential operator.
2. Feature extraction stage: To extract iris image pattern using bi-dimensional Wavelets
3. Feature matching stage: To perform the feature matching process with the XOR function applied to the iris code generated and iris template codes stored in the database.

Wildes [2] proposed the algorithm which first convert image into a binary edge map and then detect circle using Hough transform. Laplacian filter at multiple scales is used to extract features. Finally, the matching between two iris images is done using normalized correlation. S´anchez-Reillo [4] used the left and right portion of the iris in order to avoid the missing data due to eye lashes along with Gabor filters for feature extraction. Liu Yang [5] encrypted the iris code using one way coupled map lattice in order to protect of stored template data. Zhonghua Lin and Bibo Lu [7] used the imaginary coefficients of Morlet Wavelet Transform at different scales to generate the binary code of the iris image. Jing Huang et al., [8] proposed iris recognition based on non separable wavelet. After decomposing iris image into wavelet sub band coefficients using sixteen non-separable wavelet filters, Generalized Gaussian Density (GGD) modeling of each non separable orthogonal wavelet coefficient was carried for feature extraction. The Kullback Leiblar distance between GGDs was computed for matching. Mohammed Abdullah [9] presented an algorithm using wavelet transform for iris recognition where the feature vector is stored in the form of binary code.

All above mentioned techniques are computationally intensive and the size of the template database formed by the extracted feature vectors is large.

## 3. Proposed Scheme

The proposed scheme consists of following processing stages as shown in Figure 1 and the detailed procedure in case of iris identification process is as follows,

### 3.1 IRIS Recognition

The stages involved in most iris recognition systems consist of following basic modules as shown in Figure 1.

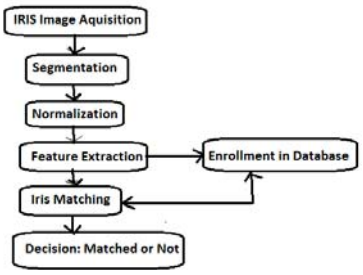



Fig.1. Typical iris recognition schemes

### 3.2 Segmentation

The segmentation module locates the position of the iris within the image by isolating it from the sclera, pupil, eyelids, and eyelashes.

#### 3.2.1. Canny Edge Detection and Localization

Canny edge detection is used to create an edge map [6, 9]. The Canny method finds edges by looking for local maxima of the gradient of the iris image. The Canny edge detects strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges. Here the boundary of an iris is located using parameters like centre coordinates x and y, the radius r, which are related according to the following equation,

$$x^2 + y^2 = r^2 \quad (1)$$

In performing the preceding edge detection step, the derivatives of the horizontal direction is to detect the eyelids, and the vertical direction is to detect the outer circular boundary of the iris. The radius of the iris image is determined and provided to the Hough transform. For better accuracy, the Hough transform is carried out initially for iris/sclera boundary and then for iris/pupil boundary.

#### 3.2.2. Hough Transform

The Hough transform is a feature extraction technique used in image analysis, computer vision, and digital image processing [2]. It finds imperfect instances of objects within a certain class of shapes by a voting procedure. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in a so-called accumulator space that is explicitly constructed by the algorithm for computing the Hough transform.

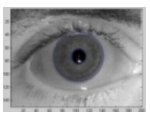


Fig.2. Detection of circular boundaries of pupil and iris.

Canny edge detection is used to build the edges in horizontal direction and then the Parabolic Hough transform is applied on it to detect the eyelids, approximating the upper and lower eyelids with parabolic arcs. If the maximum Hough space is below the threshold then it indicates the non occlusion of eyelids. For isolating eyelashes it is very easy by utilizing thresholding. This is because they are darker while comparing with further elements in eye [2].

### 3.3 Normalization

Once the segmentation module has estimated the iris's boundary, the normalization/iris unwrapping module transforms the iris texture from Cartesian to polar coordinates.

The normal Cartesian to polar transformation is recommended which maps the entire pixels in the iris area into a pair of polar coordinates (r, θ), where r and θ represents the intervals of [0 1] and [0 2π] as shown in figure 3.

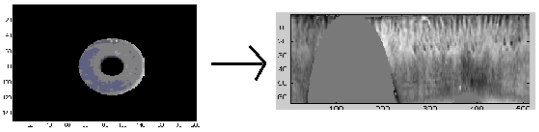


Fig.3. Normalized Iris

Normalization has advantages like, It accounts for variations in pupil size due to changes in external illumination that might influence iris size, It ensures that the irises of different individuals are mapped onto a common image domain in spite of the variations in pupil size across subjects etc.

#### 3.3.1. Histogram Equalization

Histogram equalization is done on each iris template to generate an image whose intensity also covers the entire range of intensity levels. The normalized iris image has very low contrast and it could have a non-uniform brightness in different parts of the image due to the light applied at the acquisition time. This makes the iris texture seem to be with less contrast than it really is. The contrast enhancement of the image is accomplished by means of histogram equalization in order to use the full spectrum of gray levels, hence the textures are highlighted (see figure 4). Further, filtering operation can be applied to remove noisy components.

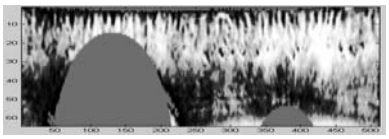


Fig.4. Enhancement of the iris normalized image.

### 3.4 Feature extraction

**Step 1:** In feature extraction stage, the above processed iris region is first resized and segmented into non-overlapping 8x8 non-overlapping blocks as shown in figure 5.

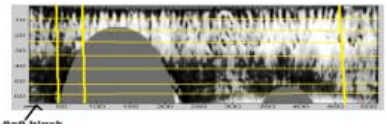


Fig.5. Normalized iris image resized in convenient form

If X denotes the rectangular iris template having size NxM, then

$$X = \left(x(1), x(2), ..., x(N_b)\right) \quad (2)$$

where, $x(i)$ is i[th] block and Nb=(NxM/64) is total number of blocks.

**Step 2:** Convert each block into 1D vector and apply Walsh Hadamard Transform (WHT) to each of the blocks as,

$$WX_k = \frac{1}{N}\left(H_M X_k\right) \quad (3)$$

and the corresponding inverse transform as,

$$X_k = \left(H_M WX_k\right) \quad (4)$$

Where the square and symmetric Hadamard transform matrix $H_m$ of order m is recursively defined as,

$$H_m = \begin{pmatrix} H_{m/2} & H_{m/2} \\ H_{m/2} & -H_{m/2} \end{pmatrix} \quad (5)$$

for m>1 and $m=2^k$ with $H_1=[1]$.

Since $H_m$ contains only the +1 or -1 entry, the transformation requires only real additions and subtractions. Further, using matrix factorization or matrix partitioning on the Hadamard matrix of order $m= 2^n$, Fast Walsh Hadamard Transform (FWHT) can be implemented. For an N-sequence, the number of operations is reduced from ($N^2$) to ($N\log_2 N$) additions and subtractions and thus. increasing the speed of processing [12-14].

**Step 3:** The energy compaction characteristics of WHT captures texture variations in its coefficients. The statistical parameters can be used to capture both local and global variations of iris texture in order to create feature vector template further. Calculate the energy of each block vector as,

$$E_k = \frac{1}{N_k}\sum_{i=1}^{N_k-1} \| WX_i \|^2 \quad (6)$$

Where, $E_k$ is energy of k[th] block and $N_k$=64 is total number of WHT coefficients in each vector.

**Step 4:** Calculate the standard deviation of each block vector as,

$$S_k = std(X_k) \quad (7)$$

Where, $S_k$ is standard deviation of $k^{th}$ block.

**Step 5:** Capture the global variations/features of iris using mean value of energy (MVE) and mean value of standard deviation (MSD) of whole iris template image which are given as,

$$MVE = \frac{1}{N_b}\sum_{l=1}^{N_b} E_l \quad (8)$$

$$MSD = \frac{1}{N_b}\sum_{l=1}^{N_b} S_l \quad (9)$$

Where, $N_b$=(NxM/64) is total number of blocks.

**Step 6:** Form binary image template using both local and global iris texture variations using the following criteria.

For a $k^{th}$ block, if both $E_k$ is greater than MVE and $S_k$ is greater than MSD then set all pixels of corresponding 8x8 block of binary template as 255 i.e. all white pixels.

Else set all pixels of corresponding 8x8 block binary template as 0 i.e. all black pixels as shown in figure 6.

Fig.6. Binary image template formed using MVE and MSD in WHT domain.

**Step 7:** Form final binary bit stream/unique code B corresponding to above binary iris image template using following rule,

If all pixels of 8x8 block is marked as 0 then corresponding bit will set as 0 else corresponding bit will set as 1.

$$B = \left(b(1), b(2), ..., b(N_b)\right) \quad (10)$$

This bit pattern will be stored in database for recognition purpose. So the size of overall database is reduced as only binary bitstream of $N_b$ bits is stored instead of NxM sized binary image template. This also increases the computational speed of searching the code during matching process.

**Step 8:** Further to increase the security of the system, the above binary bit stream B is first encrypted using the user key (K) obtained from user password and then the encrypted bit pattern template is stored.

### 3.4.1 Iris Template Matching Process

The matching algorithm consists of all the image processing steps that are carried out at the time of enrolling the encoded iris template in database. User also needs to input the same password to form user key (K). Once the bit encrypted bit pattern B' corresponding to binary image formed is extracted, it is tried to match with all stored encrypted bit patterns B using simple boolean XOR operation. The dissimilarity measure between any two iris bit patterns is computed using Hamming Distance (HD) which is given as,

$$HD = \frac{1}{N_l}\sum_{l=1}^{N_l} X_l (XOR) Y_l \quad (11)$$

Where, $N_l$=total number of bits in each bit pattern, $X_l$=bit pattern corresponding to the iris image to be matched and Yl=bit pattern stored in the template database. As HD is a fractional measure of dissimilarity with 0 representing a perfect match, a low normalized HD implies strong similarity of iris codes.

## 4. Experimental results

### 4.1 Experimental setup

In order to test the performance of the proposed method, a set of eye images obtained from the *Bath University Repository* [10] were used. Figure 7 shows an example of the eye images contained in this database. The data set consists of all grayscale images. Matlab R2009b is used on Intel Core 2 Duo (2.1 GHz) machine with 2GB RAM for the simulation purpose. The time needed for iris recognition was approximately 2-3 seconds. The searching period depends on the database size. There were total 80 numbers of irises stored in the database for this experiment. Figure 8 shows the outputs obtained at each processing stages along with the final binary image template.

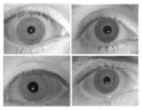
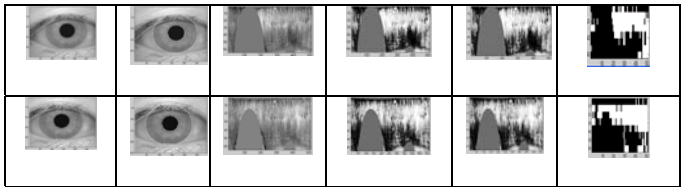


Fig.7. Examples images of the Bath University database

A unique bit stream is constructed from it as shown in figure 9, which is further encrypted using user key K and final encrypted bit stream is stored in the database. This drastically reduces size of total database along with providing higher security against compromise of template database.



| Original Images | Pupil and Iris Detec-tion | Norma-lized Iris | Histogram Equali-zation | Filtering | Binary Template |
|---|---|---|---|---|---|

Fig.8. Results obtained at different stages.

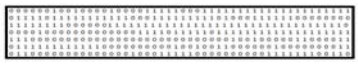


Fig.9. Sample Unique Binary Bit Pattern created representing Iris Features

### 4.2 Performance Evaluation

Following metrics are used to evaluate the performance of the system.
1) False Acceptance Rate (FAR): FAR is the measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user [11]. A system's FAR typically is stated as the ratio of the number of false acceptances divided by the total number of identification attempts.
2) False Rejection Rate (FRR): A statistic used to measure biometric performance when operating in the verification task and defined as the percentage of times the system produces a false reject. A false reject occurs when an individual is not matched to his/her own existing biometric template.
3) Equal Error Rate (EER): EER is the rates at which both accept and reject errors are equal. In general, the lower the equal error rate value, the higher the accuracy of the biometric system. The EER is sometimes referred to as the "Crossover Error Rate".
The above performance parameters are evaluated by splitting total database of 100 persons into 80 and 20 persons. The database is created with 6 images per person i.e., total number of images in the database are 480. FRR is calculated by comparing seventh image of every individual with 480 images in the database of 80 persons. FAR is calculated by considering 20 individuals as imposters and are compared with 480 images in the database. Table 1 shows the resulted FRR and FAR for the proposed and existing technique.

Table1. FAR and FRR Comparison

| User | Proposed Scheme | |
|---|---|---|
| | FAR % | FRR % |
| 1-10 | 0.01 | 0.02 |
| 10-20 | 0.05 | 0.03 |
| 20-30 | 0.02 | 0.05 |
| 30-40 | 0.08 | 0.10 |
| 40-50 | 0.02 | 0.01 |
| 50-60 | 0.04 | 0.05 |
| 60-70 | 0.11 | 0.05 |
| 70-80 | 0.03 | 0.01 |

From the result, it can be observed that the proposed technique results in lesser FRR and FAR. The value of EER obtained is 0.106. From all the results obtained, it can be said that the proposed technique results in better accuracy in recognition/verification process.

## 5. Conclusions

In this correspondence, we propose a novel robust iris recognition scheme having less computational complexity along with higher accuracy. Automatic segmentation is achieved through the use of the canny edge detector and hough transform for localising the iris and pupil regions. The present work has empirically demonstrated the good performance of statistical properties like Mean Value of Energies and Mean Value of Standard deviations in WHT domain as a robust iris object Descriptor. The energy-compaction characteristics of WHT are used to capture iris texture variations. In order to reduce the size of the database, binary bit stream instead of binary image is stored in the database for matching purpose. Fast Walsh Hadamard Transform Algorithm along with reduced feature vector size provides faster recognition rate. Further to increase the security of the system, the bit stream obtained is first encrypted using the user key obtained from user password and then the encrypted bit pattern template is stored. Experimental results show that the proposed algorithm provides lesser FRR and FAR values during matching along with less computational complexity and better security. The future work will be carried out for real applications utilization such as generation of compact iris codes for mobile phones and PDAs.

**Acknowledgments**

Author would like to thank Defence Institute of Advanced Technology, DRDO Lab, Ministry of Defence, India for providing necessary facilities to carry out research. Author also would like to thank University of Bath, Bath for making available a large iris database for our research purpose.

## References

[1] J. Daugman,"High confidence visual recognition of persons by a test of statistical independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pp. 1148–1161, 1993.
[2] R. Wildes, "Iris recognition: An emerging biometric technology," Proceedings of the IEEE, vol. 85, no. 9, pp. 1348–1363, 1997.
[3] J. Daugman, "How iris recognition works," IEEE Trans. Circuits Syst. Video Techn., vol. 14, no. 1, pp. 21–30, 2004.
[4] R. S´anchez-Reillo and C. S´anchez-´Avila, "Iris recognition with low template size," in AVBPA, ser. Lecture Notes in Computer Science, J. Big¨un and F. Smeraldi, Eds., vol. 2091. Springer, 2001, pp. 324–329.
[5] Liu Yang, Yue Xue Dong, Liu Ying Fei and He Yan, "Iris Recognition System Based on Choas Encryption," IEEE International Conference on Computer Design and Applications, vol 1, pp. 537-539, 2010.
[6] S. V. Sheela, P. A. Vijaya, "Iris Recognition Methods – Survey", International Journal of Computer Applications, 2011.
[7] Zhonghua Lin and Bibo Lu, "Iris Recognition Method Based on the Imaginary Coefficients of Morlet wavelet Transform," Seventh IEEE international Conference on Fuzzy Systems and Knowledge Discovery, pp. 573-577, September 2010.
[8] Jing Huang, Xinge You, Yuan Yan Tang, "Iris Recognition Based on Non Separable Wavelet," IEEE International

Conference on Systems, Man and Cybernetics, pp. 1552-1557, 2008

[9] Mohammed A M Abdullah, F H A Al-Dulaimi, Waleed Al-Nuaimy and Ali Al-Ataby, "Smart Card with Iris Recognition for High Security Access Environment," IEEE International Conference on Biomedical Engineering, pp. 382-385, 2011.

[10] D. Monro, "Bath University iris database," University of Bath, Bath, School of Electronic and Electrical Engineering, 2008,http://www.bath.ac.uk/elec-eng/research/sipgl.

[11] Li Ma, Tieniu Tan, Yunhong Wang, Dexin Zhang, "Personal Identification based on Iris Texture Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.12, pp. 1519 – 1533, 2003.

[12] Beauchamp, K.G., 1984, "Applications of Walsh and Related Functions: with an Introduction to Sequency Theory", Academic Press, London, 295-300.

[13] Zhihua, L. and Qishan, Z.,1983, "Ordering of Walsh functions", IEEE Transactions on Electromagnetic Compatibility, 2, 115–119.

[14] Brown, R.D., 1977, "A recursive algorithm for sequency-ordered fast Walsh transforms", IEEE Transactions on Computers, C-26, 8, 819–822.

[15] C.Anand Deva Durai, M.Karnan, "Iris Recognition Using Modified Hierarchical Phase-Based Matching (HPM) Technique", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 8, May 2010

[16] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal Identification Based on Iris Texture Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, pp. 1519-1533, 2003.

[17] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient Iris Recognition by Characterizing Key Local Variations," IEEE Transaction Image Processing, vol. 13, pp. 739-750, 2004.

# Building MultiView Analyst Profile From Multidimensional Query Logs: From Consensual to Conflicting Preferences

**Eya Ben Ahmed[1], Ahlem Nabli[2] and Faïez Gargouri[3]**

**[1] Computer Science Department, Higher Institute of Management of Tunis, Tunis, Tunisia**

**[2] Computer Science Department, Faculty of Sciences of Sfax, Sfax, Tunisia**

**[3] Computer Science Department, Higher Institute of Computer Science and Multimedia of Sfax, Sfax, Tunisia**

**Abstract**

In order to provide suitable results to the analyst needs, user preferences summarization is widely used in several domains. In this paper, we introduce a new approach for user profile construction from OLAP query logs. The key idea is to learn the user's preferences by drawing the evidence from OLAP logs. In fact, the analyst preferences are clustered into three main pools : (i) consensual or non conflicting preferences referring to same preferences for all analysts; (ii) semi-conflicting preferences corresponding to similar preferences for some analysts; (iii) conflicting preferences related to disjoint preferences for all analysts. To build generic and global model accurately describing the analyst, we enrich the obtained characteristics through including several views, namely the personal view, the professional view and the behavioral view. After that, the multiview profile extracted from multidimensional database can be annotated.

***Keywords:*** *data warehouse, text mining, clustering, profile, preferences, conflict, OLAP logs, annotation.*

## 1. Introduction

Data warehouses store a large amount of information which are analyzed in order to support strategic decision makers. OLAP analyses consist in exploring interactively the data warehouse using navigational operations. To better fit the analyst's needs, several complicated operations may be performed. Generally, some analyses are usually made by the same decision makers. Despite the diversity of the analysts' intentions, existing OLAP technology provides regularly the same results for the same keyword queries. The main reason behind this is that the search process is made out of the user features.

In fact, collecting relevant user interests and main user preferences in a user profile, may efficiently enhance support personalization and user-centric adaptivity. The notion of user profiling has been introduced in order to personalize applications so as to be tailored to the user needs. The user profile may contain different types of information: *personal data* such as identity, demographic data; *professional data* such as position/function, principal responsibilities, role and duties; and finally *behavioral data* mainly related to the data warehouse schema preferences.

Accordingly, our work focuses on the multiview analyst profile building from OLAP logs: First, we prepare the text in a preprocessing stage. Second, we cluster behavioral information in consensual, semi-conflicting and conflicting preferences. Then, we generate a generic user profile through its enrichment by mandatory views. Finally, the derived user profile may be annotated.

The rest of the paper is organized as follows: Section (2) introduces the work related to user profile modeling. In section (3), we present our approach for user profile construction and annotation. In order to validate our contribution, we describe the three steps we followed to carry out the OLAPAnalystProfile system. Experimental results evaluating the efficiency of our system are reported in section (4). In section (5) we conclude our work and briefly outline future work.

## 2. Related Work

In this section, we focus on the various research work closely related to the domain of the user profile content and the user profile modeling in data warehouse area.

Table 1: Comparison of user profile modeling approaches

| *Method* | *Area* | | *Acquisition* | | *Semantic* | | *Term* | | | *Conflict* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | IR | DW | Explicit | Implicit | Ontological | Non ontological | Atemporal | Short-term | Long-term | Conflicting | Non conflicting |
| Gowan [10] | X | | X | | | X | X | | | | X |
| Sieg et al. [16, 17] | X | | | X | X | | | X | X | | X |
| Liu et al. [9] | X | | | X | X | | X | | | | X |
| Challam et al. [3] | X | | | X | X | | X | | | | X |
| Cherniack et al. [4] | X | | | X | | X | X | | | | X |
| Bouze-ghoub, Kostadinov [2] | X | | X | | | X | | X | X | | X |
| Ravat et al. [11, 12] | | X | X | | | X | | | X | | X |
| Jerbi et al. [7, 8] | | X | X | | | X | | X | X | | X |
| Rizzi and Golfarelli [6, 13, 14] | | X | X | | | X | | | X | | X |
| **Our proposal** | | **X** | | **X** | | **X** | **X** | | | **X** | |

## 2.1 User profile content

Several definitions of user profile are proposed. According to its representation in Information Retrieval (IR) area, we distinguish the following definitions of the user profile content:

- As *weighted keyword vectors* [10]. Generally, such profile is represented using vectorial representation. For example, the user profile is composed of two keywords weighted (*status*, *query*) as follows (*i.e.* {*status*, 0.7; *query*, 0.8});
- As *semantically weighted ontological concepts* [16] combining the user's interests and Yahoo concept hierarchy.

  According to [3], the user profile is a set of weighted concepts selected from the ODP ontology. According to Liu et al. [9], the user profile consists in a set of categories and for each category, a set of terms (keywords) with weights is defined. The weight of a term in a category reflects the significance of the term on representing the user's interest in that category.
- Using *the utility notion*. In fact, the profile specification is broken into two parts [4]: the Domain clause (DOMAIN) defines and names sets of objects of interest (domain sets); and the utility clause (UTILITY) specifies the relative values of objects contained in each domain set.
- Using *several dimensions* [2], such as the user interest, the context of the launched query, the accurate level of quality, the interactions history and the different preferences on these dimensions.

## 2.2 User profile modeling in data warehouses

Taking aggregation into account, the user profile content is restricted to expressed preferences on schema rather than on instances as commonly done in Data Warehouses (DW) [1].

Ravat et al. [11,12] proposed a conceptual model of user profile based on multidimensional concepts (fact, dimension, hierarchy, measure, parameter or attribute). To assign priority weights to attributes of a multidimensional schema, the personalization rules are described using the Condition-Action formalism.

Accordingly, an OLAP query language adapted to the personalization context is proposed. The weights are taken into account during OLAP analyses. In addition, the proposed algebra contains OLAP operators allowing the drilling, rotations, selection, ordering, aggregation and modification operations.

Jerbi et al [7,8] propose a context-aware OLAP Preference model which is defined on MDB schema. Using a qualitative approach, the OLAP preferences are modeled and closely depend on user analysis context (c.f. figure 1). That's why a conceptual model of OLAP context is conceived using an arborescence of OLAP analysis elements.

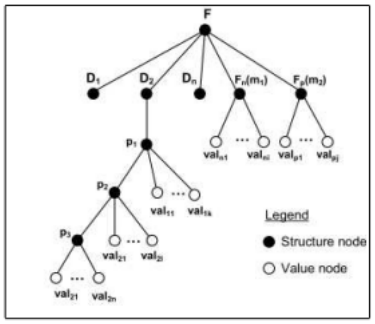



Fig. 1. OLAP analysis context Tree.

Rizzi [13] introduce MyOLAP approach where preferences are expressed using a strict partial order and are made through the first-order logic formulas.

For instance, an illustrative example of formulated preference is presented in the following. Let's consider A a set of attributes belonging to the domain *dom*(A). A

preference P is a strict partial order P = (A, < P) ∈ *dom*(A); x <P y is interpreted as *I like y better than x*.
Formulated on schema, the preferences concern not only dimensional attributes but also measures, and group-by sets. A preference algebra has been introduced to manage the different relation between preferences using specific defined operators.

The process of user profiling in data warehouse area is strictly restricted to schema personalization. Indeed, no added value information that can fundamentally orient the user preferences are taken into consideration, namely professional information such as current function, abilities and disabilities, etc. As shown by table 1, the main distinctive feature of our work is the automatic creation of analyst profile from his history in data warehouses including several views and handling the conflict aspect.

## 3. The Proposed Approach of Analyst Profile Construction And Annotation

The automatic process of OLAPAnalystProfile system is carried out in three stages (cf. Figure 2):

- ***Preprocessing***: of OLAP log queries, it consists, on the one hand, in the session and text segmentation and, on the other hand, in the identification of entities.
- ***Generic profile construction***: it consists in clustering of preferences on three main pools: (i) *consensual preferences*; (ii) *semi-conflicting* preferences and (iii) *conflicting preferences*. We propose in this stage a new similarity measurement between the preferences. After that, such preferences are enriched using mandatory information to build generic and global profile.
- ***Profile Annotation***: in order to facilitate the classification, adding information in the profile-content, correlating two preferences or investigation of future actions in the created profile, the annotation may take several forms, such as element of the data warehouse schema and its frequency.

### 3.1 Preprocessing

In the case study, the preprocessing allows the text segmentation into sentences and the delimitation of entities.

**Segmentation**: is the determination of the sessions boundaries, on the one hand, and sentences borders, on the other hand. The existing tools can be categorized as follows: (i) some of them take into account all typographical markers, (ii) other tools are backboned on linguistic bases (i.e. the syntactic structure of a sentence or the significance of each typographical marker).

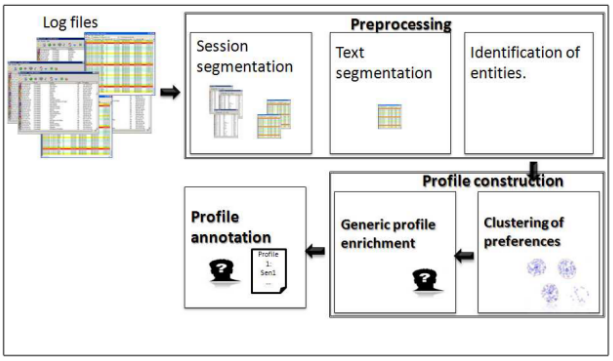



Fig. 2. Architecture of the analyst profile construction and annotation system.

Taking benefit from the structure of an MDX query, we have developed our own segmentor relying on both punctuation marks and the MDX query form. The result of the text segmentation of OLAP log file is shown by Figure 3.

```
<?xml version="1.0" encoding="UTF-8"?>
<Log session = "SalesManager1">
<Query id = "1">
<Columns id = "1"> Sales Amount </Columns>
<Rows id = "1"> Year=2010 </Rows>
<Rows id = "2"> Year=2011 </Rows>
<Cube id = "1"> Sales </Cube>
<Condition id = "1"> [Tunisia].[Tunis] </Condition>
</Query>
<!-...-
</Log session>
</xml>
```

Fig. 3. Extract of the OLAP log segmentation.

**Named Entities Recognition** : Named Entities are types of specific lexemes referring to an entity of the concrete world in given domain, namely social, medical, economic or geographical area and having a particular name [5]. The entities are identified in the log files by a tag which corresponds to the entity type. The types selected are recognized by rules using the multidimensional schema considered as a dictionary of named entities. Figure 4 presents the same text of Figure 3 after the named entities identification. Initially the position of each term is fixed. Then, each entity is recognized by specifying its attributes.

```
<?xml version="1.0" encoding="UTF-8"?>
<Log session = "SalesManager1">
<Query id = "1">
<Measure id = "1"> Sales Amount </Measure>
<Dimension id = "1"> Time
<Member id = "1"> Year=2010 </Member>
<Member id = "2"> Year=2011 </Member>
</Dimension>
<Cube id = "1"> Sales </Cube>
<Condition id = "1"> [Tunisia].[Tunis] < /Condition>
<Dimension id = "1"> Place
<Member id = "1"> Tunisia </Member>
<Member id = "2"> Tunis </Member>
</Dimension>
</Query>
<!–...-
</Log session>
</xml>
```

Fig. 4. Extract of the OLAP log segmentation after the named entities recognition.

## 3.2 Generic Profile Construction

This step is composed of clustering of OLAP queries highlighting the analyst preferences on the one side and enrichment of created profile on the other side.

**Clustering Of Preferences**: In this stage, we categorize the behavioral information extracted from the preprocessed log files on three main pools: (i) *consensual or non conflicting preferences* referring to same preferences for all analysts; (ii) *semi-conflicting preferences* corresponding to similar preferences for some analysts; (iii) *conflicting preferences* related to disjoint preferences for all analysts.

We apply the complete link hierarchical clustering algorithm to gather the queries. In fact, this algorithm merges in each step the two closest clusters having the biggest similarity distance. The latter is computed using a similarity measurement between queries.

*A. Similarity Measurement Between Queries*

A number of similarity measures is used in the hierarchical clustering to discover the closest pair of documents to merge. Among them, the cosine measure [15] is commonly the most used in document clustering particularly when the number of frequent concepts on each document is drastically different. In addition, this measure is based on the document components and is not sensitive to the document length.

We have chosen the Jaccard distance because it significantly suits the large documents. Thus, we extend this measure to the multidimensional context. The multidimensional Jaccard distance relies on the MDX query structure, particularly on similarity between used facts, measures, dimension attributes, as well as slicer specification members. The later is used in the Where clause and restricts the result data. Any dimension that does not appear on an axis in the SELECT clause can be named on the slicer. The similarity measure is the number of common facts, measures, dimensions and slicer specification members in the two queries divided by the total number of facts, measures, dimensions and slicer specification minus the already computed numerator, it is computed according to the following formula. We suppose

$(A) = C_{Fact(q_i, q_j)} + C_{Measure(q_i, q_j)} + C_{DimensionAttribute(q_i, q_j)} + C_{SSMember(q_i, q_j)}$.

$$J(q_i, q_j) = \frac{(A)}{\left[\sum_{k=i,j} Fact(q_k) + \sum_{k=i,j} Measure(q_k) + \sum_{k=i,j} DimensionAttribute(q_k) + \sum_{k=i,j} SSMember(q_k)\right] - [(A)]}$$

with $C_{Fact(q_i, q_j)}$: Common facts of $q_i$ and $q_j$,

$C_{Measure(q_i, q_j)}$ : Common measures of $q_i$ and $q_j$,

$C_{DimensionAttribute(q_i, q_j)}$ : Common dimension attributes of $q_i$ and $q_j$,

$C_{SSMember(q_i, q_j)}$ : Common slicer specification members of $q_i$ and $q_j$.

For example, we consider the two following queries $q_1$ and $q_2$.

***q1**: SELECT [Measures].[Sales Amount] ON COLUMNS, [Date].[All]ON ROWS*
*FROM Sales*
*WHERE ([Customer].[France].[Lyon]);*

***q2**: SELECT [Measures].[Sales Amount] ON COLUMNS, [Date].[2010], [Date].[2011]ON ROWS*
*FROM Sales*
*WHERE ([Customer].[France].[Lyon]);*

$q_1$ and $q_2$ use the same fact, the same measure and the same slicer specification member. However, $q_1$ uses all dimension attributes of the Date dimension which are 5 and $q_2$ accesses to only two dimension attributes which are 2010 and 2011.

We suppose $(A) = C_{Fact(q_1, q_2)} + C_{Measure(q_1, q_2)} + C_{DimensionAttribute(q_1, q_2)} + C_{SSMember(q_1, q_2)}$.

$$J(q_1, q_2) = \frac{(A)}{\left[\sum_{k=1,2} Fact(q_k) + \sum_{k=1,2} Measure(q_k) + \sum_{k=1,2} DimensionAttribute(q_k) + \sum_{k=1,2} SSMember(q_k)\right] - [(A)]}$$

$$= \frac{1+1+2+1}{2+2+2+5+2-(1+1+2+1)} = \frac{5}{13-5} = 0.625$$

*B. Hierarchical Clustering Algorithm*

The different construction steps of hierarchical clustering undertaken in our OLAPAnalystProfile are described as follows:

- *Initialization*: Count up the frequencies of each query. Let each query be a cluster; if its frequency is greater than 1, consider only a cluster for each group of repetitive queries;
- *Treatment*: Compute similarity matrix;
- *Assignment*: Merge the two closest clusters based on the two following conditions: *(a)* the maximum of similarity measure using multidimensional Jaccard distance; (b) *the maximum of frequent queries*;
- *Updating*: Update similarity matrix;

- *Iteration*: Repeat steps 3 and 4 until only three clusters remain.

A good clustering method will produce high quality clusters with high intra-cluster similarity and low inter-cluster similarity. To measure the conflict between clusters, we integrate the concept of frequency. Finally, we stop running the algorithm when the number of clusters reaches three which are: (i) *consensual or non conflicting preferences*; (ii) *semi-conflicting preferences* corresponding to similar preferences; (iii) *conflicting preferences*.

For instance, let us consider the four following queries:

*$q_1$: SELECT [Measures].[Sales Amount]ON COLUMNS, [Date].[All]ON ROWS*
*FROM Sales*
*WHERE ([Customer].[France].[Lyon])*
*$q_2$: SELECT [Measures].[Sales Amount] ON COLUMNS, [Date].[2010], [Date].[2011] ON ROWS*
*FROM Sales*
*WHERE ([Customer].[France].[Lyon])*
*$q_3$: SELECT [Measures].[Sales Amount]ON COLUMNS, [Product].[Astradol] ON ROWS*
*FROM Sales*
*$q_4$: SELECT [Measures].[Sales Amount] ON COLUMNS, [Date].[All]ON ROWS*
*FROM Sales*
*WHERE ([Customer].[France].[Lyon])*

Table 2: Similarity matrix

| Distance | $C_1$: Freq($C_1$)=2 | $C_2$: Freq($C_2$)=1 | $C_3$ : Freq($C_3$)=1 |
|---|---|---|---|
| $C_1$: Freq($C_1$)=2 | 0 | 0.625 | 0.222 |
| $C_2$: Freq($C_2$)=1 | 0.625 | 0 | 0.333 |
| $C_3$ : Freq($C_3$)=1 | 0.222 | 0.333 | 0 |

We start by counting the frequencies of the queries: (*i*) Freq($q_1$)=2; (*ii*) Freq($q_2$)=1; (*iii*) Freq($q_3$)=1; (*iv*) Freq($q_4$)=2. After that, we assign clusters to queries as follows: (*i*) $q_1 \Rightarrow C_1$ ; (*ii*) $q_2 \Rightarrow C_2$ ; (*iii*) $q_3 \Rightarrow C_3$ ; (*iv*) $q_4 \Rightarrow C_1$. In fact, the first and the forth queries are merged because they are identical. Then, we compute the similarity matrix shown by the table 2. The similar pair of queries is $C_1$ and $C_2$, at distance 0.625 and $C_1$ is the most frequent. These queries are merged into a single cluster called "$C_1/C_2$ ". Then we compute the distance from this new compound query to all other queries. In complete link clustering, the rule is that the distance from the compound query to another query is equal to the greatest similarity distance from any member of the cluster to the outside query. So the distance from "$C_1/C_2$ " to $C_3$ is chosen to be 0.333 which is the distance from $C_3$ to $C_2$ , and so on. After merging $C_1$ with $C_2$, we obtain the matrix illustrated by the table 3. The running example is a sample of our data set. However, in real case when we reach the three clusters, we may finally stop merging.

Table 3: Similarity matrix after merging $C_1$ with $C_2$.

| Distance | $C_1/C_2$ | $C_3$ |
|---|---|---|
| $C_1/C_2$ | 0 | 0.333 |
| $C_3$ | 0.333 | 0 |

**Generic Profile Enrichment:** As outlined by the conceptual modeling of analyst profile shown by figure 5, the behavioral component of the derived profile may be enriched by adding:

- *personal information* such as identity and demographic data. They include the user identity specified using his name, his social security number, etc, demographics identified using his age, his gender, his address, his marital status, his number of children, etc, his professional contacts as well as his credit card number. Generally, such kind of information does not need frequent update.
- *professional information* such as position/function (e.g. sales manager), principal responsibilities (e.g. for sales manager; to achieve the company's goals and to develop the people reporting to them), role (e.g. for the sales manager, to focus on sales; to set sales objectives, forecasting, budgeting, organizing and sales force's recruitment) and duties (e.g. for sales manager; to assign sales territories, or geographic regions to selling personnel; to evaluate the performance of the sales workers; to represent his company at trade association conventions and meetings; to promote his products, etc).

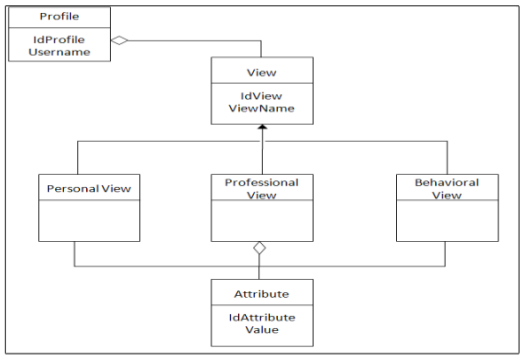



Fig. 5. Conceptual modeling of analyst profile.

## 3.3 Profile Annotation

We continue the enrichment of the profile by other metadata which will be very useful for all ulterior treatment (information retrieval, automatic summarization, storage of the preferences in a database, indexation, etc).

Mainly, we annotate the user profile-content by adding frequency to each clause of behavioral preference. Eventually, we store each preference and each related annotation in a separate database.

For instance, we present an example of profile annotation shown by the figure 6. Indeed, the cluster is annotated through the frequencies of the fact Sales, the measure Sales Amount, the dimension Date and the slicer specification members Customer.France.Lyon and Customer.France.Paris which are respectively 2, 2, 2, 1 and 1.

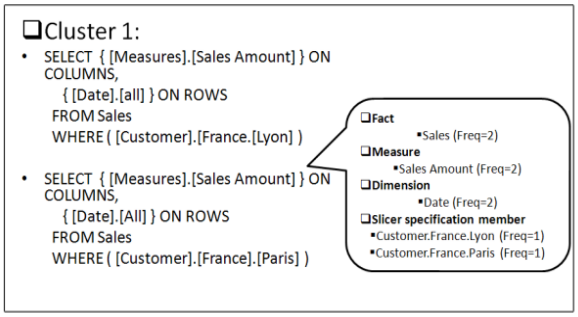



Fig. 6. Generic profile annotation.

## 4. Experimental results

In order to validate our approach, we have implemented our system OLAPAnalystProfile using the java language. In fact, our system contains three modules:

- *Module one*: Preprocessing through sessions segmentations and entities identification in log files;
- *Module two*: Generic profile construction through the clustering of preferences on three pools: (i) *consensual preferences*; (ii) *semi-conflicting preferences*; (iii) *conflicting preferences*; then its enrichment in order to derive an extended profile;
- *Module three*: Annotation of generated profile.

We prepared a corpus of 5000 OLAP queries stored in log file. We used the Weka 3.6.5 edition to apply the decision trees. First, we segment them in sessions then in queries in order to process the log files. Second, we identify the named entities based on the data warehouse schema. Then, we start the construction of profiles through the clustering of queries on three pools: (i) consensual preferences; (ii) semi-conflicting preferences; (iii) conflicting preferences. Hence, a hierarchical clustering algorithm is applied and an innovative extension of Jaccard measure is proposed in the multidimensional context. Then, an enrichment of the generated profile is performed through adding personal and professional information to behavioral ones. Finally, such a profile may be annotated using the frequency of each clause of behavioral preferences.

Our training set is presented by a part of group of queries being the output of the preprocessing step. Such a set is annotated by an expert. For each OLAP query, the default value of the preference attribute is ”conflicting”, the expert may change this value and affect the ”semi-conflicting” and ”consensual” values.

An ARFF file (the input file of Weka) is generated for each OLAP log file. It is used as an input for the used classification algorithms, namely, *ID3* which is a decision tree method based on the computation of entropy to generate the information gain and select attributes, *Classification Via Clustering* which is a simple meta - classifier that uses a cluster for classification. For cluster algorithms that use a fixed number of clusters, like SimpleKMeans, the user has to make sure that the number of clusters to generate are the same as the number of class labels in the dataset in order to obtain a useful model., *Multi class Classifier* which is a classification method handles multi-class datasets with 2-class distribution classifiers, *Hyperpipes* which is a classification algorithm constructed for each category; it contains all points of that category (essentially records the attribute bounds observed for each category); the test instances are classified according to the category that ”most contains the instance”, and *CVParameterSelection* which is a classification algorithm for performing parameter selection by cross-validation for any classifier.

Aiming to evaluate our clustering method, we apply the metrics usually of use:

- The **True Positive (TP)** rate is the proportion of examples which were classified as class x, among all examples which truly have class x, i.e. how much part of the class was captured. It is equivalent to **Recall**;
- The False Positive (FP) rate is the proportion of examples which were classified as class x, but belong to a different class, among all examples which are not of class x;
- The **Precision** is the proportion of the examples which truly have class x among all those which were classified as class x;
- The **F-Measure** is simply (2∗Precision∗Recall)/(Precision+Recall) , a combined measure for precision and recall;
- The **Receiver operating characteristic** (ROC) is the relationship between the TP and FP rates.

Table 4: Results of classification with ten-fold cross validation

| Classification Method | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC | Preference Class |
|---|---|---|---|---|---|---|---|
| ID3 | 1 | 0.004 | 0.996 | 1 | 0.998 | 0.996 | Conflicting |
| | 0.994 | 0 | 1 | 0.994 | 0.997 | 0.995 | Semi-Conflicting |
| | 0.996 | 0.001 | 0.998 | 0.996 | 0.997 | 0.997 | Consensual |
| Classification via clustering | 1 | 0.304 | 0.765 | 1 | 0.867 | 0.848 | Conflicting |
| | 0.595 | 0.1 | 0.666 | 0.595 | 0.628 | 0.747 | Semi-conflicting |
| | 0.398 | 0.034 | 0.797 | 0.398 | 0.531 | 0.682 | Consensual |
| Multi class classifier | 1 | 1 | 0.498 | 1 | 0.665 | 0.5 | Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.499 | Semi-Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.499 | Consensual |
| Hyper pipes | 1 | 0.453 | 0.687 | 1 | 0.814 | 0.773 | Conflicting |
| | 0.994 | 0.033 | 1 | 0.909 | 0.994 | 0.997 | Semi-Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.7 | Consensual |

Table 5: Results of classification with twenty-fold cross validation

| Classification Method | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC | Preference Class |
|---|---|---|---|---|---|---|---|
| ID3 | 1 | 0.004 | 0.996 | 1 | 0.998 | 0.996 | Conflicting |
| | 0.994 | 0 | 1 | 0.994 | 0.997 | 0.995 | Semi-Conflicting |
| | 0.996 | 0.001 | 0.998 | 0.996 | 0.997 | 0.997 | consensual |
| Classification via clustering | 1 | 0.353 | 0.765 | 1 | 0.849 | 0.823 | Conflicting |
| | 0.349 | 0.017 | 0.666 | 0.349 | 0.499 | 0.666 | Semi-conflicting |
| | 0.648 | 0.083 | 0.797 | 0.648 | 0.683 | 0.782 | Consensual |
| Multi class classifier | 1 | 1 | 0.498 | 1 | 0.665 | 0.498 | Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.499 | Semi-Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.497 | consensual |
| Hyper pipes | 1 | 0.478 | 0.675 | 1 | 0.806 | 0.761 | Conflicting |
| | 0.994 | 0.017 | 0.953 | 0.994 | 0.973 | 0.997 | Semi-Conflicting |
| | 0 | 0 | 0 | 0 | 0 | 0.683 | consensual |

In our experiments, as far as the value of the cross validation fold increases, the evaluation criteria produce better results and our preferences are correctly classified as illustrated by table 5. For evaluation of an error rate, we used the both of ten-fold cross validation and twenty-fold cross validation : all cases were randomly re-ordered, and then the set of all cases is divided into respectively ten and twenty mutually disjoint subsets of approximately equal size.

To assess the performance of our method, several classification methods were launched. As shown by the table 4, the ID3 algorithm engenders a precision equal to 99.6% for the first class, 100% for the second class and 99.8% for the third class. However, the classification via clustering technique generates a precision equal to 76.5 % for the first class, 66.6% for the second class and 79.7% for the third class. While MulticlassClassifier algorithm produces a precision equal to 49.8% only for the first class. Finally, Hyperpipes brings a precision equal to 68.5 % for the first class and 100% for the second class. Consequently, we stress out the accuracy of our proposed clustering method.

## 5. Conclusion

In this paper, we have proposed three stages to build and annotate analyst profile from OLAP log files starting, in a first stage, by the preprocessing of the log file which allows the text segmentation and the recognition of named entities. In a second stage, based on the conflict aspect, clustering of behavioral preferences in: (i) *consensual preferences*; (ii) *semi-conflicting preferences* and (iii) *conflicting preferences*. Then, enrichment of such behavioral preferences by adding personal and professional information. Finally, we may annotate the user profile-content using frequency.

There are different perspectives opened by this study. We think it would be interesting to confront the created profile and the spotted behavior. Moreover, we intend to investigate practical expressiveness of the derived user model. Finally, we plan to extend our contribution to personalize the query model.

## References

[1]. E. Ben Ahmed, A. Nabli, F. Gargouri, "A Survey of User-Centric Data Warehouses: From Personalization to Recommendation", The International Journal of Database Management Systems (IJDMS), May 2011, Volume 3, Number 2, 2011.

[2]. M. Bouzeghoub, D. Kostadinov, "Personnalisation de l'information: aperu de l'état de l'art et définition d'un modèle flexible de profils", In Proceedings of COnférence en Recherche d'Information et Applications (CORIA'05), pp. 201-218 , 2005.

[3]. V. Challam, S. Gauch, A. Chandramouli, "Contextual Search Using Ontology-Based User Profiles", In Proceedings of RIAO 2007, Pittsburgh USA, 2007.

[4]. M. Cherniack, E. Galvez, M. Franklin, S. Zdonik, "Profile-Driven Cache Management", In Proceedings of the 19th International Conference on Data Engineering, Bangalore, India , 2003.

[5]. O. Ferret, B. Grau, M. Hurault-Plantet, G. Illouz, C. Jacquemin, L. Monceaux, I. Robba, A. Vilnat, "How NLP Can Improve Question Answering", In Revue Knowledge Organization, 2002.

[6]. M. Golfarelli, S. Rizzi, "Expressing OLAP Preferences", Proceedings of the 21st International Conference on Scientific and Statistical Database Management, pp. 83-91, 2009.

[7]. H. Jerbi, F. Ravat, O. Teste, G. Zurfluh, "Management of context-aware preferences in multidimensional databases", International Conference on Digital Information Management (ICDIM'08),pp 669-675, 2008.

[8]. H. Jerbi, F. Ravat, O. Teste, G. Zurfluh, "Personnalisation du contenu des bases de données multidimensionnelles", Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA'10), Djerba, Tunisie, pp. 520, 2010.

[9]. F. Liu, C. Yu, W. Meng, "Personalized Web Search For Improving Retrieval Effectiveness", IEEE Transactions on Knowledge and Data Engineering, vol. 16, n1, pp. 28-40, 2004.

[10]. J. P. Mc Gowan, "A multiple model approach to personalized information access, Master Thesis in Computer Science, Faculty of science, University College Dublin, 2003.

[11]. F. Ravat, O. Teste, "Personalization and OLAP Databases, Annals of Information Systems", New Trends in Data Warehousing and Data Analysis, Vol. 3, pp. 7192, 2008.

[12]. F. Ravat, O. Teste, G. Zurfluh, " Personnalisation de bases de données multidimensionnelles, INFORSID, pp. 121-136, 2007.

[13]. S. Rizzi, "OLAP preferences: a research agenda", International Workshop on Data Warehousing and OLAP (DOLAP07), pp. 99-100, 2007.

[14]. S. Rizzi, "New Frontiers in Business Intelligence: Distribution and Personalization", Advances in Databases and Information Systems (ADBIS'10), pp. 23-30, 2010.

[15]. G. Salton, "Automatic Text Processing", Addison-Wesley Publishing Company, 1988.

[16]. A. Sieg, B. Mobasher, R. Burke, G. Prabu, S. Lytinen, "Representing user information context with ontologies", uahci05, 2005.

[17]. A. Sieg, B. Mobasher, S. Lytinen, R. Burke, "Using Concept Hierarchies to Enhance User Queries in Web-based Information Retrieval", Artificial Intelligence and Applications (AIA), 2004.

**Eya Ben Ahmed** is carrying out a PhD degree in Computer Sciences in Sfax University in Tunisia. Her research is focused on data mining techniques and data warehouses.

**Ahlem Nabli** obtained her Ph.D. in Computer Science from Sfax University in Tunisia in 2010. She is currently an assistant professor of Computer Science at the Faculty of Sciences of Sfax in Tunisia. Her research interests include data warehouses and ontologies.

**Faïez Gargouri** obtained his Ph.D. in Renes Descartes, Paris 5 in 1995. He is actually a Professor of Computer Science at the Higher Institute of Computer Science and Multimedia of Sfax in Tunisia. His research interests include Information system, ontology engineering, semantic web, advanced databases, and data warehouses.

# Ultra-Wide-Band Microstrip Concentric Annular Ring Antenna for Wireless Communications

**S. Azzaz-Rahmani[1], N. Boukli-Hacene[2]**

**Telecommunication Laboratory, Faculty of Technology,**
**Abou-Bekr Belkaid University**
**Tlemcen, 13000, Algeria**

**Abstract**

In this paper, a new design technique for bandwidth enhancement of concentric microstrip annular ring slot antennas is presented. Using this technique, an Ultra-Wide-Band antenna is designed with simulated bandwidth of 111.29%.

***Keywords:*** *Microstrip antenna, wideband, concentric patch, bandwidth, low impedance.*

## 1. Introduction

Microstrip patch antennas are widely used because of their several advantages such as light weight, low volume, low fabrication cost, and capability of dual, triple and several frequency operations. However microstrip antennas suffer from a number of disadvantages, particularly the narrow bandwidth[1]. This is a serious limitation of these microstrip patch antennas. Different techniques are used to overcome this narrow bandwidth limitation. These techniques include increasing the thickness of the dielectric substrate, decreasing dielectric constant and using parasitic patches [2]. These techniques have limitations like, excitation of surface waves and increase in antenna size [3].

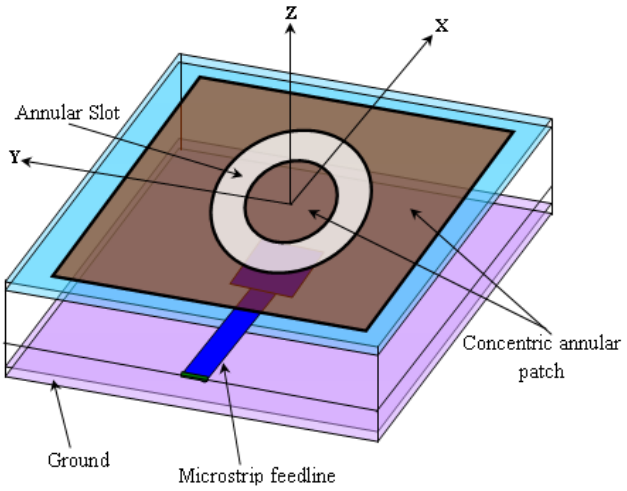
Annular ring slot antennas are considered to be among the narrowband resonant antennas [4]. Multi-element concentric ring slots have been used to design multi-band antennas. However, because of transmission zeros that exist between the different resonances, these resonances cannot easily be merged to obtain a wideband response [4, 5]. The purpose of this paper is to propose a microstrip structure which will increase the bandwidth without increasing its physical dimensions.

## 2. Antenna design

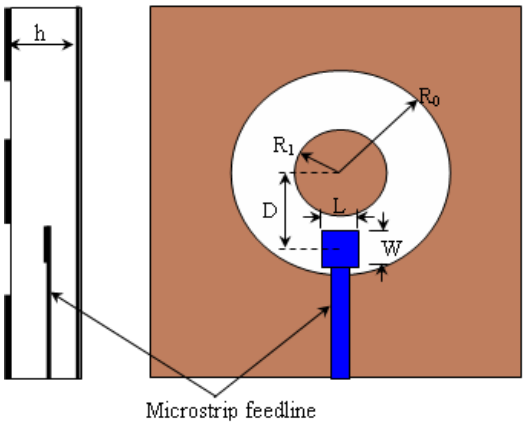
Annular ring slot antenna has a reduce size more than circular patch antenna and the ultra-wideband characteristic [6]. In this paper, to broaden the bandwidth of annular ring slot antenna, we placed the concentric annular patch inside circular slot and designed the low impedance feed line.

Because, for an annular ring slot antenna, the resonant frequency of the lowest order mode TM11 can be much lower than a circular patch of the same size, the annular ring slot antenna can be designed to the smaller size than the circular patch antenna [5, 6]. This fact could be appreciated physically by noting that the average path length travelled by the current in the annular ring is much longer than the circular ring for the lowest order mode [5, 6].

Fig. 1 shows the configuration of the ultra-wide-band concentric annular ring microstrip antenna. We placed a microstrip feed line to the bottom of a substrate with relative permittivity of 4.3 and thickness of 2mm. The concentric circular patch embedded in an annular slot is placed on the substrate to match the impedance.




(a)




(c) (b)

Fig 1. (a) Antenna structure, (b) top view, (c) side view, geometric parameters: $R_0$=26 mm, $R_1$= 10 mm, L=12 mm, W=10.5 mm and D= 18.34 mm.

The geometry parameters of our proposed concentric annular ring antenna are; $R_0$=26 mm, $R_1$= 10 mm, L=12 mm W=10.5 mm and D= 18.34 mm.

## 3. Simulation results

The variation of the return loss magnitude and phase of the concentric annular ring antenna as a function of frequency are shown in fig. 2 and 3 respectively. The bandwidth is calculated using the formula:

$$BW = \left[\left(1/f_c\right)\times\left(f_H - f_L\right)\right]\times 100\% \quad (1)$$

Where, $f_H$ and $f_L$ are the higher and lower frequency band respectively, for which the return loss $S_{11}$ is less than -10dB and $f_c$ is the centre frequency of this band.

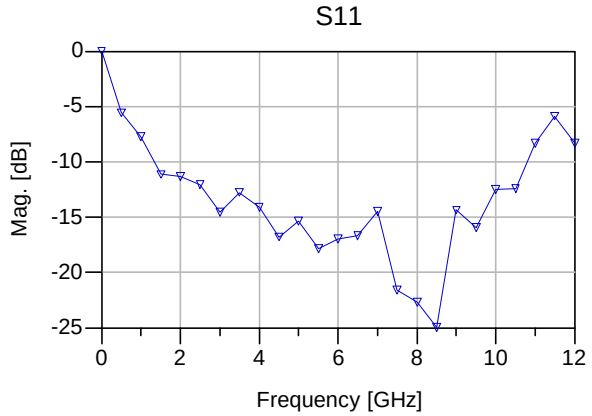



Fig 2. Simulated return loss of concentric annular ring antenna as a function of frequency.

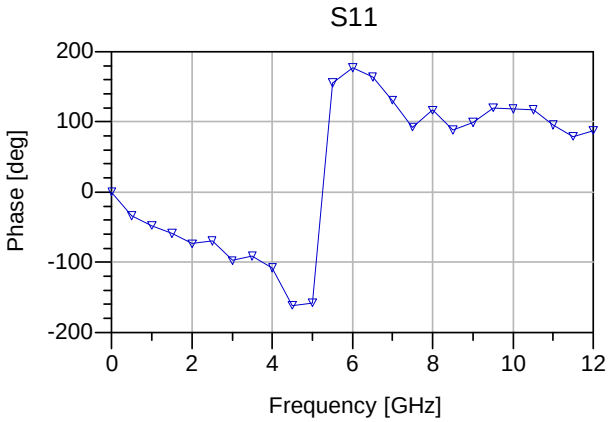



Fig 3. Simulated phase of concentric annular ring antenna as a function of frequency.

From Fig. 2, we see that the antenna operates from 1.34 to 10.8 GHz which provides a bandwidth of 111.29%.

The simulated E-plane radiation pattern is presented in Fig 4. The cross-polarization component (E_cross) is also illustrated. This pattern is simulated at 8.5 GHz.

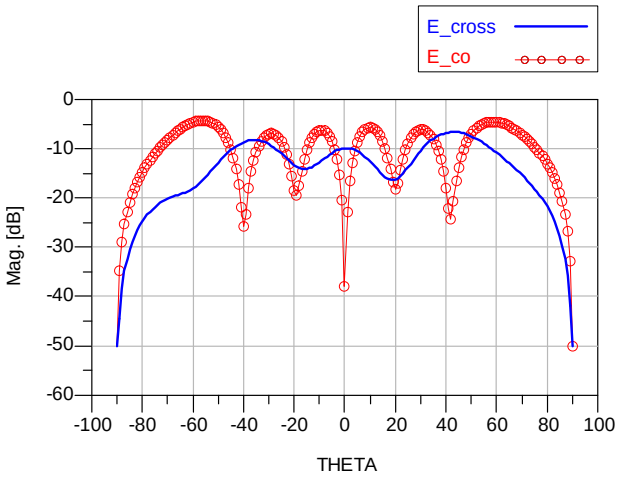



Fig 4. E-plane radiation patterns at 8.5 GHz. Simulated copolarization (E_co); simulated cross-polarization (E_cross).

Fig 5 shows 3D radiation pattern of this antenna measured at frequency of 8.5 GHz. On this plot appear several side-lobes, these shows very well the multibandes functioning.

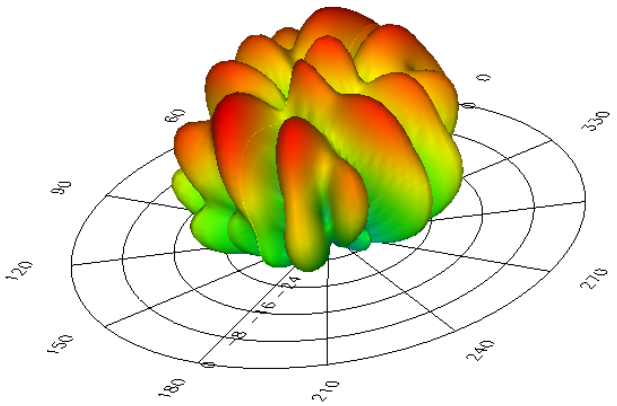


Fig 5. 3D radiation patterns measured at frequency of 8.5 GHz.

Fig. 6 shows the variation of the simulated return loss for different values of the distance between the center of circular slot and the center of low impedance feed line. It is observed that the return loss of high frequency is varied much smaller than that of low frequency.

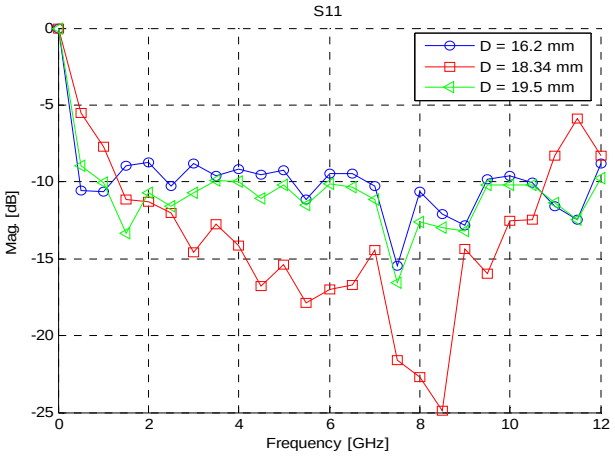



Fig 6. Variation of simulated return loss for diferent values of D.

It was also found that the distance between the center of low impedance feed line and the center of circular slot (D) have much influence on the return loss. In fact, the antenna has multiband operating: the frequency bands are respectively: [7GHz – 9.5GHz], [10.5GHz – 11.8GHz] for D=16.2mm, and [1GHz – 3.5GHz], [4GHz – 10GHz], [10.5GHz– 12GHz] for D=19.5mm. For D = 18.34 mm the antenna has an Ultra Wide Band operating where the bandwidth is 111.29%

## 4. Conclusions

A new technique for bandwidth enhancement of concentric annular ring antennas is presented. The use of a discontinuous microstrip feed line has permitted to obtain an antenna bandwidth equal to 111.29%, which is much larger than that of a conventional ring antenna.

Using this technique, we obtained an Ultra-wide bandwidth with small size antenna. It may find proper applications in wideband mobile communication system.

## References

[1] Ramesh Garg, Prakash Bartia, Inder Bhal and Apsiak Ittipiboon, “Microstrip Antenna Design Hand Book,” Artech House, Norwood, MA, 2001.

[2] D.M.Pozzar “Microstrip Antenna Coupled to Microstripline,” Electron Lett., Vol. 21, No.2, pp. 49-50, January 1995.

[3] Y. Coulibaly and T. A. Denidni, "Design of a Broadband Hybrid Dielectric Resonator Antenna for X-Band Applications," Journal of Electromagnetic Waves and Applica-tions, Vol. 20, No. 12, pp. 1629-1642. 2006.

[4] N. Behdad and K. Sarabandi, “Wideband double-element ring slot antenna” Eelectronics Letters, Vol. 40 No. 7 , pp. 408 – 409, April 2004,

[5] H.K. Kan, R.B. Waterhouse and D. Pavlickovski, “Compact dual concentric ring printed antennas” IEE Proc.-Microw. Antennas Propag., Vol. 151, No. 1, pp. 37-42, February 2004.

[6] Debatosh Guha, , Sujoy Biswas, Manotosh Biswas, Jawad Y. Siddiqui,, and Yahia M. M. Antar, Fellow, IEEE, “Concentric Ring-Shaped Defected Ground Structures for Microstrip Applications », IEEE antennas and wireless propagation letters, Vol. 5, pp.402-405, Dec 2006.

**Salima Azzaz-Rahmani** was born in Algeria in 1981. She obtained here engineering degree in 2003 and a magister degree from Abou Bekr Belkaid University, (Tlemcen) Algeria, in 2006. She is a doctorate student in the same university. Currently she is a lecturer at Djillali Liabes University (Sidi Bel-Abess). Here research interests are the analysis and syntheses of microstrip concentric annular ring and ultra wideband antennas.

**Noureddine Boukli-Hacene** Noureddine Boukli-Hacene was born in 1959 in Tlemcen, Algeria. He received the ‘Diplome d’Etudes Approfondies’ in microwave engineering (DEA Communications Optiques et Microondes) and the Doctorate degree (prepared at the Centre National d'Etudes Spatiales, Toulouse, France) in electrical engineering from Limoges University, France, in 1982 and 1985 respectively. Recently, he is a Lecturer at the University of Tlemcen. His research interests include, among others, microstrip antennas and microwave circuits.

# Adaptation of learning resources based on the MBTI theory of psychological types

**Amel Behaz[1], Mahieddine Djoudi [2]**

**[1] Faculty of Science, Batna University, code (05000) Algeria**

**[2] Laboratory XLIM-SIC and TechNE a Research Group, UFR Sciences SP2MI, University of Poitiers**
**Teleport 2, Boulevard Marie et Pierre Curie BP 30179 86962 Futuroscope, Chasseneuil Cedex- France Country**

**Abstract**

Today, the resources available on the web increases significantly. The motivation for the dissemination of knowledge and their acquisition by learners is central to learning. However, learners show differences between the ways of learning that suits them best.

The objective of the work presented in this paper is to study how it is possible to integrate models from cognitive theories and ontologies for the adaptation of educational resources. The goal is to provide the system capabilities to conduct reasoning on descriptions obtained in order to automatically adapt the resources to the learner according to his preferences. We rely on the model MBTI (Myers-Briggs Type Indicator) for the consideration of learning styles of learners as a criterion for adaptation.

***Keywords:*** *Learner modeling,* learning style*, MBTI,* adaptive learning*, knowledge engineering, semantic web, ontology.*

## 1. Introduction

The use of information technology and communication has greatly improved the way we read and learn. These advances are revolutionizing our way of learning by facilitating access to content and services. A large amount of educational resources is produced continuously on the Web. Given the cost of production of these resources and the expertise to produce them, it is essential to make them easily accessible, usable and reusable.

Students can learn, communicate and collaborate by means of Learning Management Systems (LMS) such as Blackboard [1], Moodle [2] or ATutor [3]. The problem is that LMS doesn't offer personalized services, all the students being given access to the same set of educational resources and tools, without taking into account the differences in knowledge level, interests, motivations and goals.

In fact, we are aware that any adaptation process is based on a model of users. The adaptation necessitates enough knowledge of the users (capacity, objectives, learning preferences, history). People have different personalities, which affect their daily activities, emotions, the ways they interact, and how they learn. In the initial stage of the research, we intended to generate a framework to understand how we could review the effectiveness of the adaptive hypermedia systems. Identifying each student's learning preferences style was considered. To deal with this aspect, Myers-Briggs Type Indicator (MBTI) questionnaire test was employed. The MBTI test was originally developed to measure people's personalities type [4]; however, it has also been used for developing different teaching methods that meet different students' learning styles. Of course, MBTI cannot be used to stigmatize each person as a particular personality type, but it has demonstrated many useful tips to improve learners' communication style with tutors through constructive use of differences [5]. In particular, the education domain has used it to develop different teaching methods that meet different students' learning styles.

Our aim is to try to apply this model in the learning context and to study the advantages of learners' styles of learning as a criterion for adaptation.

In this research paper, we will introduce our new approach of modeling that based on ontologies. A learner's ontology based on the theory MBTI. Our aim, here, is to describe and analyze the preferential needs of learning process. Besides, a simple ontology related to the field of knowledge supported with resources is put forwards. This is carefully structured using the concepts and relations. Finally, the adaptation model will be carefully elaborated by applying norms of adaptation selection and, also,

presentation. To support our proposal, we will, ultimately present a prototype and a conclusion of our whole work.

## 2. Learning style

Learning style can be defined as "attitudes and behaviors that determine the preferred way of a person to learn" [6]. We can say that learning style is the way a person perceives and organizes information. An overview of the literature quickly show the plurality and diversity of learning style models. These models are grouped into three types:
- Models of learning style preferences are interested in the conditions of teaching and learning.
- Models of learning style with an interest in how the learner processes information in terms of preferred means
- Models of learning styles that address the personality of the learner. Example: Myers and Briggs [7].

As part of our approach, we examined the model Myers-Briggs Type Indicator (MBTI).

### 2.1 MBTI model

The MBTI is based on the work of Carl Gustav Jung and the authors of the instrument, Isabel Briggs Myers and her mother, Katharine Cook Briggs. [8]. His work led in turn to Myers Briggs Type Indicator or MBTI. The MBTI is a tool that allows an individual to be aware of his own behavioral preferences. According to this theory, each has a natural preference. When a person uses his favorite pole, it generally succeeds better and feel more competent, natural and dynamic.
This indicator was used for many years in the Anglo-Saxon countries, including the army and in schools to lead students to the university that will fit best their profile [9]
The MBTI is based on the principle that the differences in behavior from one person to another can be expressed in terms of preferences between the polarities. Four bipolar oppositions thus define four main dimensions of psychic life:
**I**: Introversion (Introvert) **E**: Extraversion (Extrovert)
The scale E / I shows preference to direct his attention to the outer world of people and things (E) or to the inner world of ideas (I).
**S**: Sensation (Sensing) **N**: intuition (intuitive)
The scale S / N indicates the preference of the perception of things, events or details of the present moment (S) or the possibilities, the intuitions of the future (N).
**S**: Sensation (Sensing) **N**: intuition (intuitive)
The scale S / N indicates the preference of the perception of things, events or details of the present moment (S) or the possibilities, the intuitions of the future (N).
**T**: thinking (Thinking) **F**: feeling (Feeling)
The scale T / F indicates the preference of the rational decision is based on an objective analysis and logic (T) or on subjective values (F).
**J**: Judgement (Judging) **P**: Perception (perceiving)
The scale J / P indicates the preference for the organization and control of external events (J) or for the observation and understanding of these events (P).
The various combinations of these preferences result in a total of 16 personality types and are typically denoted by four letters to represent a person's tendencies on the four scales as shown in Table 1.

Table 1: The 16 MBTI types

| | | | |
|---|---|---|---|
| ISTJ | ISFJ | INFJ | INTJ |
| ISTP | ISFP | INFP | INTP |
| ESTP | ESFP | ENFP | ENTP |
| ESTJ | ESFJ | ENFJ | ENTJ |

For example, ENFP stands for Extraversion, iNtuition, Feeling, and Perception. This does not mean that the person has only four preferences, but that the four preferences show a greater presence than their counterparts. There are questionnaires to determine the personality type of a person.

## 3. Learner model

The focus of our research is on the learning style as the adaptation criterion, since it is one of the individual differences that play an important role in learning, according to educational psychologists. Learning style refers to the individual manner in which a person approaches a learning task. Research in this area began relatively recently and only a few systems that attempt to adapt learning styles have been developed.
Several works have proposed solutions based on ontologies to describe learner's profile. [10], [11], [12] have developed models for the representation of learners in order to monitor and control their activities.
Our modeling approach is based on the findings of works on cognitive theory for the description of learner's profile and more specifically for the representation of styles (preferences for learning).
We suggested that the learner's model is defined as an ontology comprising the diverse qualities and characteristics of the user according to special concepts and relations between them. We would like to introduce a learner's description according to four facets (see Fig. 1). These latters are considered as abstract notions, in our ontology.

The first facet called "**Identity**" is used to represent information about a particular learner. It is composed of predefined attributes that are essential and common to all users: name, surname, login, language, media type ... and it is modeled as a set of attribute-value pairs.

The second facet called "**Preferences**" predefined attributes that are prerequisite and common for all learning preferences. This component is directly based on the theory of psychological types of MBTI (Myers-Briggs Type Indicator). According to this theory, everyone has a natural preference. We have also used it for developing different teaching methods that meet different students' learning styles. The individuals show differences in the ways of learning as follows:

- Some prefer basic, complete instruction: (T);
- Some prefer to start directly learning about the task: (F);
- Some prefer to get through the subject first before the following: (J);
- Some need flexibility, opportunities for exploration: (P);
- Some need space and time to learn (L);
- Finally, some assimilate faster in learning: (R)

This component is modeled as a conceptual vector Vp= ($P_I$, $P_E$, $P_S$ $P_N$, $P_T$, $P_F$, $P_J$, $P_P$) this vector enables us to specify MBTI the characteristics of the psychological style of the learner and, hence, find out more about his preferences in the learning process. There are questionnaires which determine the psychological type of a person. For example, the types of psychological types of a learner A1 are described as follows: Vp= (I:2%, **E**:18%, **S**:20%, N:6%, **T**:25%, F:5%, **J**:20%, P:4%). A learner A1 has the type **ESTJ** so he is a learner who prefers to finish his task before moving on to the next aspect that emphasizes the well-structured, etc.

The third facet which is referred to as "**Capacity**" is introduced to represent and measure the degree of a learner's understanding of a concept. This perception is represented by a stereotype (class or category of individuals) this can be achieved with "quiz" test. The stereotype model eases the categorization of perception in a given group: the learner is classified with a particular category, acquiring his specificity and adaptability feature provided by the mentioned stereotype. The possible observations are: very low, low, medium, good, excellent. This scale of evaluation provides more precision.

The fourth facet called "**History**" is supposed to record everything about the learner memorization of navigation and resources in the documents read. This recording gives exact information like: the length of a resource or the navigational course.

Both the third and the fourth aspects develop automatically and dynamically while the learner is acquiring new concepts.

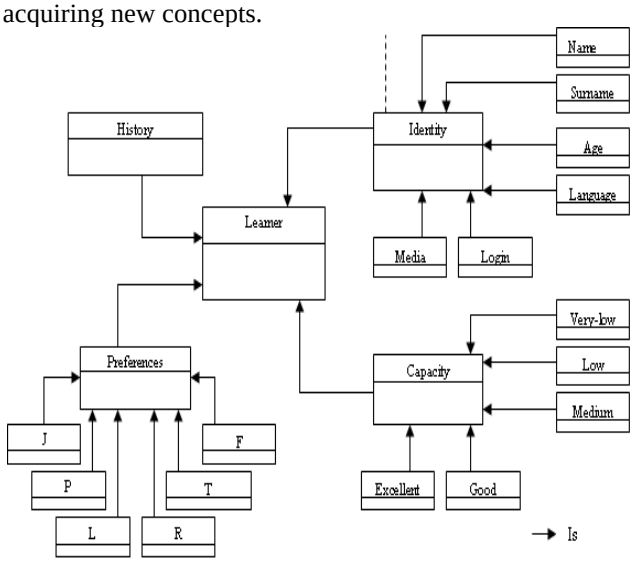



Fig. 1 Learner ontology.

## 4. Domain model

Works in semantic web [13], [14] and [15] Endeavor to render the content of resources accessible by using the adaptive hypermedia systems. The semantic web uses "engineering knowledge" as effective means of representation of knowledge. The main characteristics of the structure of the "semantic web" (common meaning, metadata processed by machines) seem very effective to resolve the problem of searching for pertinent information.

In this context, a pedagogical resource is pedagogical, atomic unity representing a physical entity (text, picture, sound...) belonging to a given category (definition, example, illustration exercise....) corresponding to a particular notion. These resources are represented as XML fragments, carefully arranged together to form hypermedia pages. To represent a pedagogical resource, we can take into consideration different information. Information about the resource itself (Norm LOM) [16], information about the notion covered by the resource, knowledge of the resource category, and information about the learning style of the resource. All this knowledge is represented by using ontologies. The aim is to add semantic annotations to the pedagogical resource content to make them easily found and adapt them to the different learners' profiles.

Many facets are suggested for the description of a resource. A complete ontological model containing different representation aspects is illustrated in (see Fig. 2).

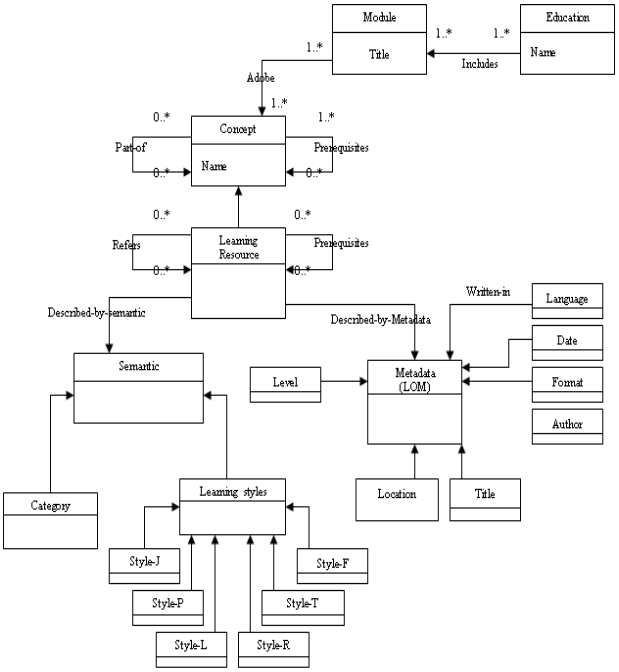



Fig. 2. Complete ontological model for representing a learning resource

The "**Thematic**" facet helps to represent the pedagogical resources according to the themes they deal with in the form of different modules. Each module covers one or many notions. A particular notion of the domain (data base, network...) represented with a name, is an abstract representation of a finite set of pedagogical resource. These concepts to be exploited are described with a graph in which the nodes are the concepts and the arcs are the semantic relationships between them. This graph is conceived by the system administrator. It is obvious that this task is not easy in case the domain covered is vast. However, there exist similar ontologies (not forcibly with the same semantic relations).The different type's definite relations between concepts are:

- Prerequisite: concept X is the prerequisite of the concept Y if the learning of Y requires the knowledge of X.
- Part of: X part of Y if X is a concept belonging to Y.

The relations between the concepts influence the adaptation of the hypermedia, for example, certain resources cannot be added to a page because their concepts have prerequisites that have not yet been acquired by the learner. The link "Part of" is, also, very useful for the adaptation. We can, for example, divide a concept into simple concepts and introduce less complex pedagogical resources to the learner that are adapted to his knowledge. Choosing a course then a module among many via a list of choice is done by choosing and visualizing the "OWL" ontology that describes the concepts of a particular domain. This enables a learner to specify his knowledge (already acquired, and to be acquired) relative to a domain.

The second facet "**Metadata LOM**" ensures the description of a pedagogical resource by using Metadata (author, title, date, language, media, location.....) this part is similar to Metadata described in LOM norm. But to meet the needs of our application, and to make the analysis easier, we have established a set of vocabulary. Let's, for example, consider R1= <language, {"French"}>, <media, {"text", "video", "picture"}>, <author, {"behaz", "djoudi"}>

A LOM description is attached to each pedagogical resource. This description is illustrated in figure 2. The associated Metadata provides précis's and well classifies information about each learning resource rendering the ulterior researches more effective.

The different types of relations between the resources are:

- Prerequisites: if the reading of the resource A necessitates the reading of the resource B: "an exercise" requires the knowledge of a "definition" first.
- Cite: the resource A cites the B if A contains a reference or a link towards B.

Or, the representation suggested in the norm is not sufficient for accessibility and adaptation in a hypermedia system. We complete it with a semantic representation of contents.

The third facet "**Category**" helps to classify the pedagogical resources under different categories (introduction, example, definition, illustration, exercise....) depending on their contents.

The fourth facet "**Learning styles**" enables to take into account the different learning styles of a pedagogical resource. This facet, which helps to specify the content of a resource, is adapted to a learning style seen as psychological MBTI. For example, the resource R1 has the conceptual vector Vs = (I: 4%, **E**: 16%, **S**: 22%, N: 4% **T**: 20%, F: 10%, **J**: 18%, P: 6%), indicates that this object is most suitable for a MBTI type **ESTJ** indicates that this resource is more adapted to the profile of a leaner who prefers to finish his task before going on to the following one (style-T: 20%) that prefers the well structured aspect (style-J: 18%) which assimilates long (style-S: 22%). The operation of the resource parameters is ensured by the designer (or an annotator) about the content and the possible usage. This operation is realised via an ergonomic interface (forms, questionnaires.....etc.) hiding the

technical details during the creation of a pedagogical resource. After that, the value of a usage vector Vs of a resource can be modified (or adjusted) manually by the designer of the resource, or automatically by the system. This modification is based on the ulterior traces of usage of that resource through the different learners' profiles.

## 5. Adaptation model

The adaptation model is used to generate personalized content from the information space of the model learning and applying the rules of structure and presentation. There are many approaches to model adaptation using logic Woukeu [17] Stash [18], which are often based on the use of rules. In our case, after identifying the learning styles of learners. The adaptation process takes place as it follows:

- Research : Identification of resources to the concept
- Selection: the choice of adequate resources in the best model of the learner (media favorite, level, style, ...).

### 5.1 Research of resources

When the learner has defined the concept on which it wants to work, There will be a process of research resources. The identification of resources relevant to a concept is based on the learner model and domain model knowledge. The proposed annotation resources systematically connects resources to their concepts. The system builds a request to the resource base. This returns the resource identifiers corresponding to the concept. To improve the process of finding an inference engine is built. This is mainly based on semantic links between the learning resources in the ontology and the inference rules.

### 5.2 Resource Selection

The result of the previous step is a list of resources found explicitly or implicitly inferred. This list is subject to another module that compares the semantic usage of each resource in the list with a description of the learning preferences of the learner.
This comparison is performed for using a distance D which calculates the distance (as defined vector), it returns a semantic and a measure of geometric distance. Given the two vectors, Vp= ($P_I$, $P_E$, $P_S$ $P_N$, $P_T$, $P_{F,}$ $P_J$, $P_P$) describing the learning preferences of the learner $A_i$ and the vector Vs= ($S_I$, $S_E$, $S_S$, $S_N$, $S_T$, $S_F$, $S_J$, $S_P$) describing the styles of the resource. The measure of similarity is the calculation of distance between vectors (the vector of preferences of the learner Vp and the vector of styles Vs of educational resources found in the previous step) Various measures can be used. We have used the cosine measure.

$$D = Cos\ (Vp_{Ai}, Vs_{Re}) = \frac{\sum_{i=1}^{8} P_i \sum_{i=1}^{8} S_i}{\sqrt{\sum_{i=1}^{8} P_i^2 \sum_{i=1}^{8} S_i^2}} \quad (1)$$

Another list of resources is then proposed to the learner. This set is closest to the learning preferences of the learner. This therefore ensures better assimilation of knowledge and capabilities to the rhythm of understanding of each learner.

## 6. Implementation

We implemented the system as a Web application. For this we used the Java and Servlets that allow great flexibility and portability of the application. It is within the scope of the new generation of Web (Semantic Web). In fact we used the OWL ontology to represent the developed and the Jena API for handling. Figure 3 shows the software architecture of the system developed.
A prototype is still at an experimental stage at the University of Batna. We confirm the large number of educational resources involved to qualify a system that actually adapts to the learner. An important advantage is the fine descriptions of available resources, we facilitated access. Also, the inclusion of styles (preferences) learning as a criterion of adaptation has facilitated the acquisition of knowledge.

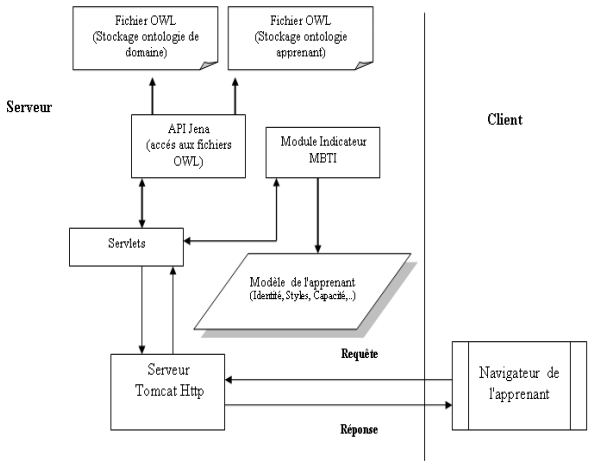



Fig. 3 Software architecture of the system developed.

When registering a new learner, a questionnaire is proposed to determine the psychological type (see Figure 4). Once completed and validated, the system calculates and stores the result in the learner profile.

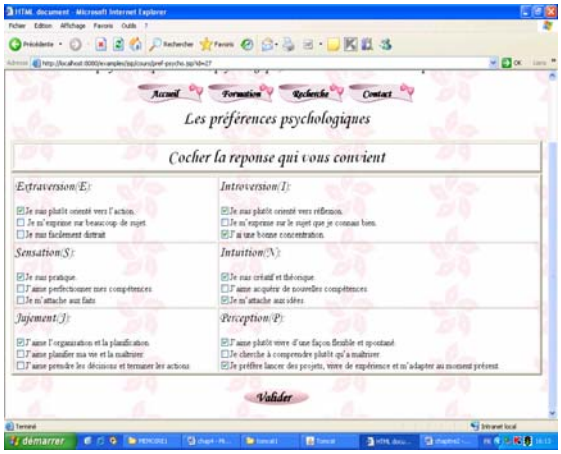



Fig. 4 Indicator MBTI.

A learner, connects to the system via an interface to describe the request. Choose a module is to choose and view the OWL ontology that describes the concepts of a particular domain.

Figure 5 shows a description of a given resource. In the left side you can see the concepts of a particular area here it is the field of computer module "Network Architecture". In the right side can see the descriptions for a given resource (LOM metadata) extracted from the ontology.

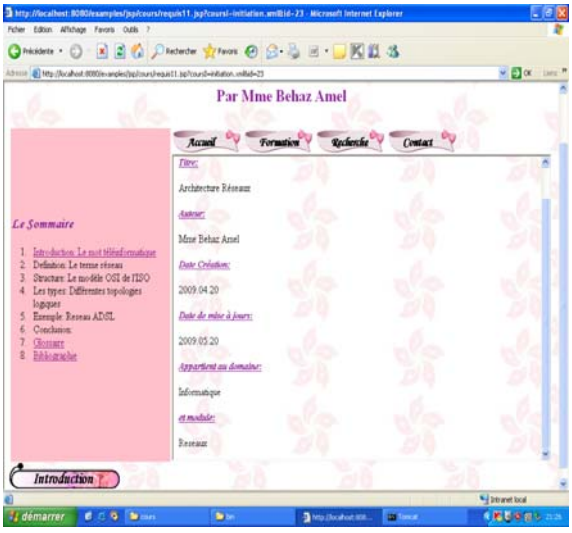



Fig. 5 Description of the resource "Architecture Networks"

In Figures (6 and 7) a dynamic generation of the same concept "Network Architecture" presented to two learners of different profiles. A learner of a medium level of knowledge that prefers to complete its task before moving on to the next, which favors the appearance well structured (style-J) another learner of excellent level needs flexibility, opportunities for exploration (style-P). We note the content which is generated differently, more details to a medium learner and less details for an excellent student.


Fig.6 Resource for a Medium learner and style -J

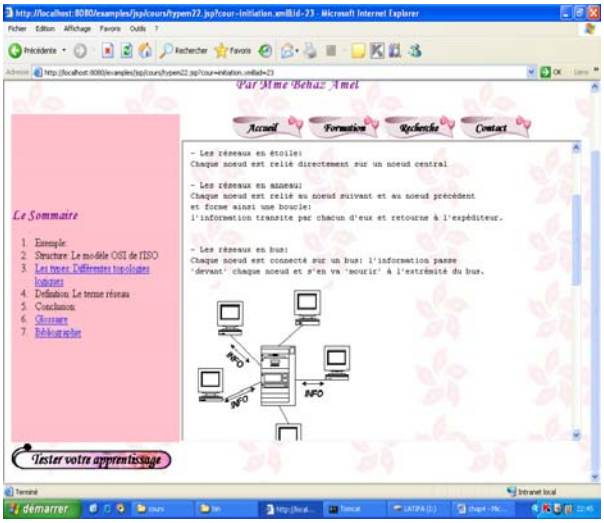



Fig.7 Resource for a Excellent learner and style -P

## 4. Conclusions

Personalized e-learning implementation is recognized one of the most interesting research areas in the distance web-based education. We conducted a research on the effects

of student's psychology to improve their learning performance.

We have introduced a new modeling adaptive system for the e-learning. A learning ontology based on the findings of works on cognitive theories for the description of learner's profiles, specifically for the representation of learning preferences. A simple domain ontology covered by resources. Finally, the adaptation model is described in using the research and selection resources.

We introduce the resources with metadata to detect them easily. We use indexation techniques of pedagogical resources. a prototype was developed; it represents the advantages of our project: easing the acquisitions of concepts by introducing the resources compatible with the learner's profiles.

We do not claim to have solved the problem of learning style modeling and adaptation. We do however hope to have shed light on some aspects and filled in some of the gaps. Further research is of course needed to clarify the remaining and newly raised issues.

## References

[1] Blackboard. Available at: http://www.blackboard.com. (visited on date 2010)

[2] Moodle Available at: http://moodle.org. (visited on date 2010)

[3] ATutor. Available at: http://www.atutor.ca. (visited on date 2010)

[4] Myers, I. Guide to the Development and Use of the Myers-Briggs Type Indicator. CPP, Inc, 3rd edition, 1998.

[5] Carolyn, S., & al., Myers Briggs Type Preferences in Distance Learning Education. International Journal of Educational Technology, 2(2) 2001.

[6] P. Honey and A. Mumford.: The Manual of Learning Styles. Maidenhead, Berkshire (1992)].

[7] Myers I. B., Mccauley, M., Quenk, N. and Hammer, A. 'MBTI Manual : A Guide to Development and Use of the Use of Myers-Briggs Type Indicator' Consulting Psychologist ress, Palo Alto (1998).

[8] C. Bishop-Clark, D. Wheeler: The Myers-Briggs Personality Type and Its Relationship to computer programming. JREC. 26:358-370 (1994)

[9] J. K DiTiberio: Uses of type in education. In MBTI Manual: A guide to the development and use of the Myers-Briggs Type Indicator, eds. I. B. Myers, M. H. McC., And N. Q. CPP (1998)

[10]Razmerita, L."User modeling and personalization of the Knowledge Management Systems", Chapter Book, in Adaptable and Adaptive Hypermedia 2005, by Idea Group Publishing.

[11]Snae C., Brueckner, M., «Ontology-Driven E-Learning System Based on Roles and Activities for the Learning Environment". Interdisciplinary Journal of Knowledge and Learning Objects, Volume 3, 1-17 2007.

[12]Zhuhadar L., Nasraoui O., and Wyatt R. "Dual representation of the semantic user profile for personalized web search in an evolving domain," in Proceedings of the AAAI Spring Symposium on Social Semantic Web, Where Web 2.0 meets Web 3.0 pp. 84- 89, 2009.

[13]Corby Querying the Semantic Web with the Corese conception & development search Engine,, proceeding of European Conference on Artificial Intelligence ECAI, 2004.

[14]Duitama F., Defude B., Bouzeghoub A., and Carpentier C. "A framework for the generation of adaptive courses based on semantic metadata", Multimedia Tools and Applications, 2005.

[15]Abi Chahine C., Kotowicz J-P., Chaignaud N., Pécuchet J-P., "Conception d'un outil d'aide à l'indexation de ressources pédagogiques": EIAH09 Environnements Informatiques pour l'Apprentissage Humain, Le Mans 2009.

[16]IEEE LOM. Available at: http://ltsc.ieee.org/wg1(visited on date 2010)

[17]Woukeu A., Wills G., Conole G., Carr L., Kampa S., and Hall W. "Ontological hypermedia in education": A framework for building web-based educational portals. Proceedings of ED-MEDIA, 2003.

[18]Stash N, Cristea A, and Paul De Bra."Explicit intelligence in adaptive hypermedia: Generic adaptation languages for learning preferences and styles". International Workshop on Combining Intelligent and Adaptive Hypermedia Methods/Techniques in Web-based Education Systems, 2005.

[19]Behaz, A., Djoudi, M. "Modélisation ontologique pour la création d'un hypermédia adaptatif". *International Journal of Information Sciences for Decision Making (ISDM), ISSN : 1265-499X*, Vol. 39, Mai (2009).

[20]Djoudi M., "eLEARNING IN ALGERIA: Experiences On E-Learning in Algerian Universities". , e-LEARNING Pratices, Cases on Challenges Facing E-Learning and National Development, ISBN 978-975-98, Vol. 1, Anadolu University, Eskisehir-Turkey, 1-31 (2009).

**Amel Behaz** received a Master in Computer Science from the University of Batna, Algeria, in 2004. She is currently a Professor at the University of Batna, Algeria. She is a member of (Adaptive Hypermedia in E-learning) research group. She is currently pursuing his doctoral thesis research on the modeling of an adaptive educational hypermedia system. Her current research interest is in E-Learning, Knowledge Engineering, Semantic Web, Ontology, and Learner Modeling. Her teaching interests include Programming, Data Bases, and Web Technology.

**Mahieddine Djoudi** received a PhD in Computer Science from the University of Nancy, France, in 1991. He is currently an Associate Professor at the University of Poitiers, France. He is a member of SIC (Signal, Images and Communications) Research laboratory. He is also a member of IRMA E-learning research group. His PhD thesis research was in Continuous Speech Recognition. His current research interest is in E-Learning, Mobile Learning, Computer Supported Cooperative Work and Information Literacy. His teaching interests include Programming, Data Bases, Artificial Intelligence and Information & Communication Technology. He started and is involved in many research projects which include many researchers from different Algerian universities.

# Province Based Design and Simulation of Indonesian Education Grid Topology

**Heru Suhartanto[1], Ivo B. Nugroho[2] and Anisa Herdiani[3]**

**[1] Faculty of Computer Science, Universitas Indonesia**
**Depok, Indonesia**

**[2] Accenture Indonesia**
**Jakarta, Indonesia**

**[3] Faculty of Information Technology, Universitas YARSI**
**Jakarta, Indonesia**

**Abstract**

This paper discusses the design and simulation of an e-learning computer network topology, based on Grid computing technology, for Indonesian schools called the Indonesian Education Grid (IndoEdu-Grid). The grid technology proposed to solve infrastructure problems faced by Indonesian ICT Network (Jardiknas).

In previous study, we designed the topology which based on two scenarios: region based and island based topology. Each scenario run in the simulator using two packet scheduling algorithms, one will be FIFO (First In First Out) Scheduler and the other SCFQ (Self-Clocked Fair Queuing) Scheduler.

In this paper we proposed a different scenario which based on province. The simulation treatments are the same with the two previous scenarios.

The simulation results showed that when using FIFO algorithm, the province based scenario has the best performance compared to Region Based and Island Based. However, this scenario is not competitive with the others when using SCFQ algorithm which is due to higher packet lifetime.

***Keywords:*** *e-learning, grid computing, IndoEdu-Grid, province-based scenario, region-based scenario, island-based scenario.*

## 1. Introduction

E-learning as a trend has been developed so rapidly that many educational organizations and institutions in numerous countries, including Indonesia adopts it. An example of an infrastructure that can be used for e-learning in Indonesia is Jardiknas( *Jejaring Pendidikan Nasional* or Indonesian ICT Network).

Jardiknas is a national-scaled WAN (Wide Area Network) that facilitates educational activities in Indonesia. This network consists of Institutional/official Jardiknas that serves online data transaction between educational institutions, College Jardiknas–Indonesian Higher Education Networks (INHERENT)–that serves science and technology research and development, School Jardiknas that serves information and e-learning accesses in schools, and Teacher and Student Jardiknas that serves personal information and e-learning accesses (Pustekkom,2009) .

Jardiknas is established with a main purpose to serve administration in central National Department of Education (*Departemen Pendidikan Nasional*) and many domestic or foreign related work units and to serves learning processes in primary, junior high, and senior high schools based on information and communication technology.

Jardiknas covers the whole areas of Indonesia, but it still has some problems as follows.

1. The current network has not been equipped with capabilities to facilitate huge data processing and cannot maximize the distributed potential resource in the whole areas of Indonesia.
2. The number of students and educational institutions that require accesses to Jardiknas will increase every year, so that the needs of wider and easier-to-access network have risen.
3. Educational activities are advancing, where the education subjects (teachers, students, and instructors) will interact with each other, share data, and perform complex calculations and simulations, such as mathematics, physics, or biomolecular models.

To solve these problems, we propose the usage of Grid computing technology for Indonesian Education

Networks–called IndoEdu-Grid based on the structure of the provinces in the country. Grid computing (or Grid) is a system that can provide resources sharing among organizations. Grid infrastructure will provide a capability to connect the resources dynamically as an ensemble to support large-scale, resource-intensive, and distributed applications (Berman, et al, 2003).

## 2. Related Research

Nugroho (2010) in Design and Simulation of Indonesian Education Grid Topology using Gridsim Toolkit discusses the design and simulation of an e-learning computer network topology, based on Grid computing technology, for Indonesian schools called the Indonesian Education Grid (IndoEdu-Grid).

The simulation, which is built using GridSim toolkit, handle two conditions or scenarios that have different network topologies based on their routers and links configuration. The first scenario is region-based topology, and the second scenario is island-based topology.

Each scenario runs in the simulator using two packet scheduling algorithms, one will be FIFO (First In First Out) Scheduler and the other SCFQ (Self-Clocked Fair Queuing) Scheduler. The processing time of the job's packets were evaluated to determine the most effective network topology.

The simulation result showed that if SCFQ scheduling algorithm is used, then the most appropriate network topology is the topology which allows its packets with the same priorities to have less possibility to collide one another in one router or link, which is the first scenario.

The simulation results also showed that SCFQ scheduling algorithm can reduce the packets lifetime at routers that have very crowded traffics. This fact implies to the decrease of the whole job processing time.

## 2. GridSim Toolkit

In this study we use GridSim toolkit which is a Java-based simulator and supported with some additional libraries. GridSim is an open-source application and licensed under GPL license, thus it encloses its source codes in its distribution package.

GridSim's rationale is that creating a testbed infrastructure for Grid system is expensive and time-consuming, even an existing testbed infrastructure is also limited in size to a few resources and domains, and testing scheduling algorithms for scalability and adaptability, and evaluating scheduler performance for various applications and resource scenarios is harder and impossible to trace (Buyya and Murshed,2000, Buyya and Murshed, 2002, Sulistio et al, 2005)

## 2. Method

The design of IndoEdu-Grid consists of four steps : create the entities, designing class diagram, designing topological scenario, and define assumptions in the simulation.

**a. The Entities**

There are three entities that build IndoEdu-Grid (Nugroho, 2010):

1. Resources
   Resource entities are responsible to perform computation on job entities in form of Gridlets sent by one or more users and send it back to the user.
2. Users
   Users are entities responsible to submit jobs in form of Gridlet objects to the resources.
3. Jobs (gridlets)
   Jobs in GridSim are represented as the objects of the class Gridlet provided by GridSim.

**b. The Class Diagram**

This simulation consists of four main classes, namely class Main, class Islands, class NetUser, and class Randomizer and several other supporting classes available in Java's default library and the GridSim's additional library, such as `ArrayList`, `Router`, `GridSim`, `GridResource`, and `GridletList` (Nugroho, 2010). Our class diagram is shown in Fig 1.

**c. The Topological Scenario**

The scenario is a representation of our thought that divides the whole territory of Indonesia directly into 31 province units. Unlike the Region Based and Island Based scenarios in (Nugroho, 2010), in this scenario, the whole territory of Indonesia is seen made up of 31 province units. Fig. 2 shows the network topology for the Province Based scenario.

**d. Assumptions in the Simulation**

There are some assumptions during the simulation and these are as follows :

1. User location was determined with the router connected to it, e.g. Jateng_User_519 user will be connected to jatengRouter. Thus, the user's geographic position is not simulated.
2. Users select the resources randomly.
3. Users send Gridlets at the same time. Same time in this case means all objects of the NetUser class execute the body method together.

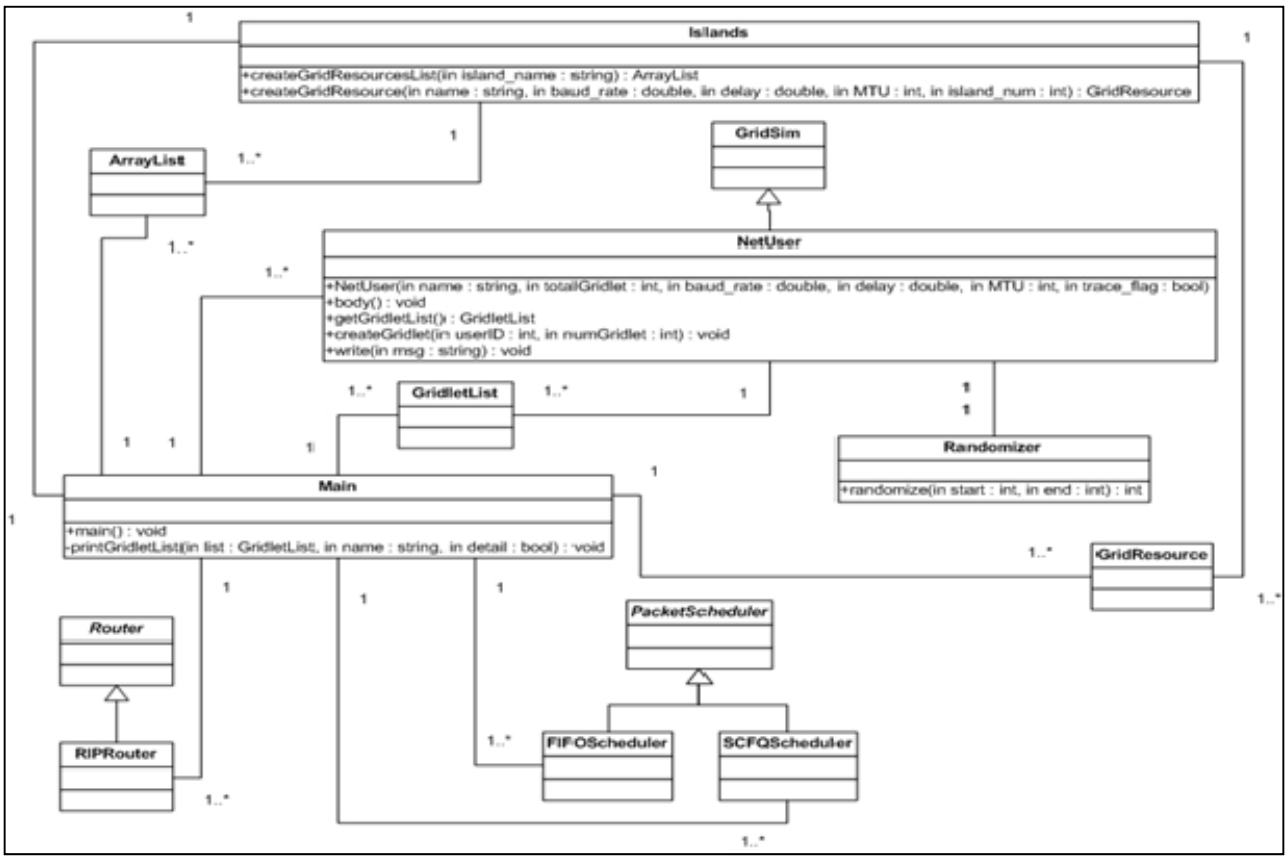



Fig. 1 The Class Diagram for Simulation Program

4. Leaf routers, edge routers, and core routers are only distinguished by entities connected to them and the configuration of the lithe configuration of the links. Leaf routers are connected to the users and GridResource entities and have links with the smallest baud rate. Edge routers are connected to leaf routers and core routers and have links with baud rate greater than the links at the leaf routers. Core routers are connected to edge routers in its area and the other core routers and have links with baud rate greater than the link at the leaf routers. Thus, there is no additional characteristic that distinguishes the functionality of these routers. These three types of router are the objects of the RIPRouter class.
5. The simulation scenarios use FIFO and SCFQ scheduling algorithm to schedule packets on the network.
6. All the resources use time-shared scheduling system.
7. Processing time is measured by obtaining the difference between the sending and receiving time of a Gridlet. Thus, the processing time includes the propagation time of Gridlet packets in the network and execution time of Gridlets. Processing time will be more influenced by the propagation time of packets on the network because the execution time of all Gridlets is the same (all resources have the same specifications of PE). Information was obtained from the CSV files created after the simulation finished.
8. Characteristics of resources–its name, its operating system, its architecture–are static properties of the resource and have no impact in the simulation.

The load of processors used to perform the simulations is ranging from 1% to 10% with the amount of free physical memory is about 43% to 50%.

**e. The computing environment.**

The simulation is run under Intel® Core™ 2 Duo T5800 processor with 2.0 GHz clock speed, 800 MHz FSB (Front Side Bus), and 2 MB L2 cache 2048 MB RAM (Random Access Memory) with shared dynamically with Mobile Intel® Graphics Media Accelerator 4500MHD; and 320 GB Fujitsu MHZ2320BH G2 SATA harddisk with 5400 rpm rotation speed. While the software are 32-bit Microsoft Windows Vista™ Business operating

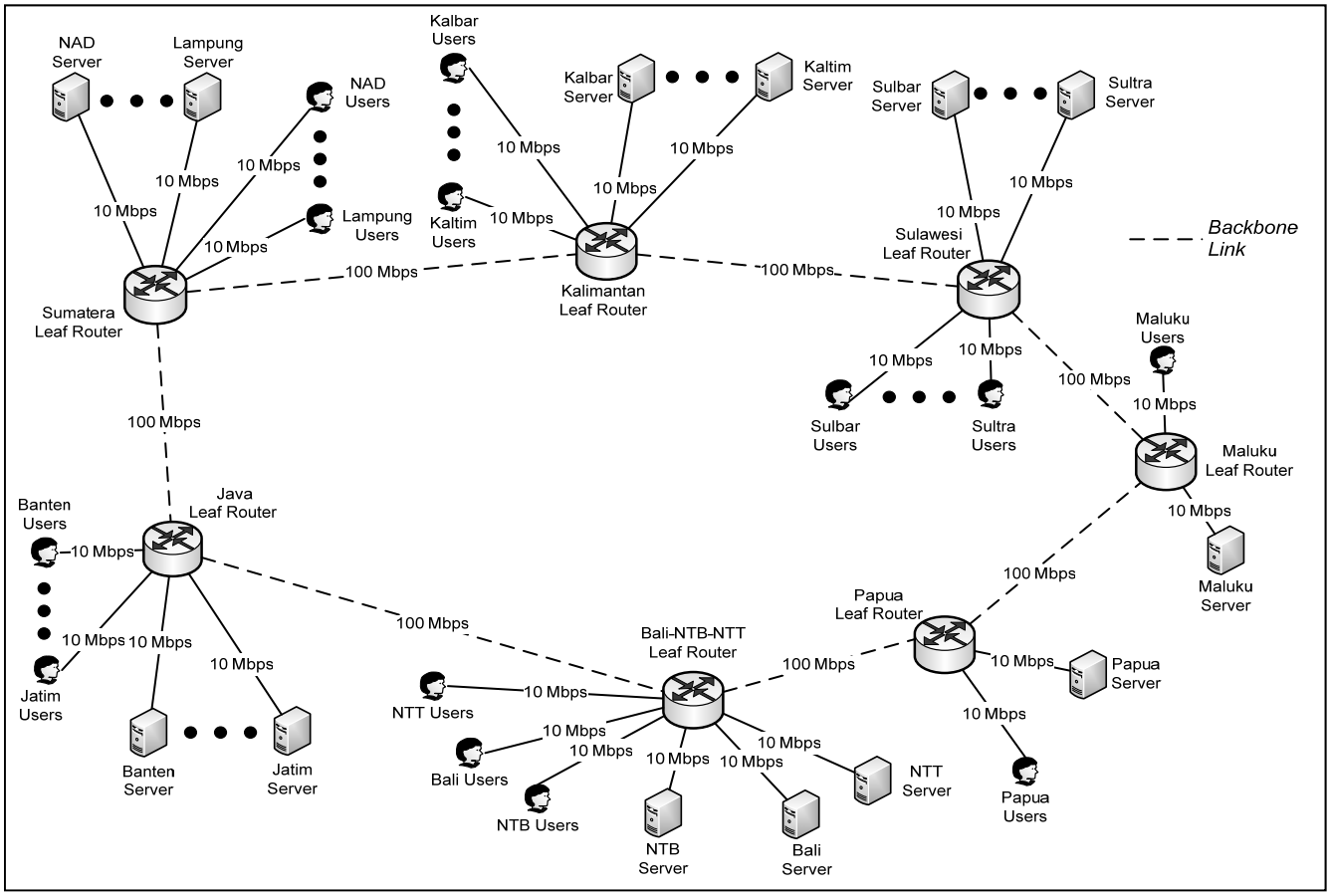



Fig. 2 Network Topology for the Provice-based Scenario

system. JDK (Java Development Kit) version 1.6.0_05 with Java™ Runtime Environment 1.6.0_05-b13. GridSim version 5.0 beta.

## 4. Result & Discussion

The results of simulation describe average processing time of the three types of Gridlets sent to the resource through province based network topologies. The results are then compared to the region based and island based topology obtained by Nugroho (2010). The simulation will show which topology that produces the lowest and highest average processing time using FIFO and SCFQ algorithm for scheduling packets. The simulation was run 10 times in each scenario to increase the validity of simulation results, and then the results were averaged.

The simulation results data per Gridlet which is the average of all provinces using the FIFO and SCFQ algorithm are shown in Table 1, while the average of all data is shown in Table 2.

Table 1: Average Simulation Results Data for the Entire Provinces per Gridlet Using FIFO and SCFQ Scheduling Algorithm

| Scheduling Algorithm | Scenario | Processing Time (in Simulation Seconds) | | |
|---|---|---|---|---|
| | | Gridlet#0 | Gridlet#1 | Gridlet#2 |
| FIFO | Region Based | 239.76471 | 184.89620 | 124.45739 |
| | Island Based | 240.23045 | 185.26774 | 124.11812 |
| | Province Based | 237.60240 | 183.48528 | 123.93875 |
| SCFQ | Region Based | 235.50311 | 180.73233 | 124.67395 |
| | Island Based | 235.78695 | 181.59782 | 124.05540 |
| | Province Based | 237.12839 | 183.29814 | 123.81650 |

From Table 2, we find that Region Based with SCFQ algorithm gives the best performance (180.30313 simulation seconds), while Island Based with FIFO gives the worst performance (183.20544 simulation seconds).

Table 2: Average Processing Time for the Entire Provinces and Gridlets Using FIFO and SCFQ Scheduling Algorithm

| Scenario | Processing Time (in Simulation Seconds) | |
|---|---|---|
| | FIFO | SCFQ |
| Region Based | 183.03943 | **180.30313** |
| Island Based | **183.20544** | 180.48006 |
| Province Based | 181.67547 | 181.41435 |

In Province Based scenario with FIFO algorithm, it appears that the processing time is small enough, i.e. 181.67547 simulation seconds. This processing time is 0,75% lower compared to the processing time of region based scenario, dan 0,84% lower compared to the processing time of island based scenario. This happened because the inexistence of edge router makes the maximum number of hops is 6 hops. By observing Figure 2, this fact can be explained by the following examples. If a user in Lampung wants to send jobs to a resource in Papua, job packets will be propagated through channels as follow :

*Lampung_User → Sumatra Leaf Router → Java Leaf Router → Bali-NTB-NTT Leaf Router → Papua Leaf Router → Papua_Res (5 hops)*

or

*Lampung_User → Sumatera Leaf Router → Kalimantan Leaf Router → Sulawesi Leaf Router → Maluku Leaf Router → Papua Leaf Router → Papua_Res (6 hops)*

In addition, the packets just have to be queued at leaf routers, so the overall processing time becomes much lower.

In Province Based scenario with SCFQ algorithm, it appears that the processing time is the largest compared to the rest two scenarios : 0,61% higher than Region Based scenario and 0,51% higher than Island Based scenario. This happened because the additional (overhead) computation done by SCFQ algorithm to select the packets to be processed first. In addition, the packets that have normal priority will have greater chance to collide with other packets that have the same priority as the available routers are only leaf router. In the end, these packets will wait and the performance of the SCFQ algorithm will approach the performance of FIFO algorithm.

Based on the results, it can be concluded that Province Based scenario has the best performance compared to Region Based and Island Based with FIFO algorithm. However, this scenario has worse performance compared to the one using SCFQ algorithm, because packet lifetime is higher.

## 4. Conclusions

In our work, IndoEdu-Grid network simulation using GridSim toolkit has been conducted. Simulation was conducted with three different types of topologies in terms of link and router configurations and two types of scheduling algorithms–FIFO and SCFQ. The final result of the average processing time is compared to obtain the most effective topology for each scheduling algorithm, while the final results of the packet lifetime is used to analyze the phenomenon happened in the simulation.

The conclusions can be drawn from the results of this simulation are as follows.

1. If the FIFO scheduling algorithm will be used in establishing IndoEdu-Grid network, then the most appropriate topology is the topology that allows the packets to have a low number of hops. In this simulation, the network with the lowest number of hops is provided by the Province Based scenario. The use of topology in the region based and island based scenario only makes the processing time becomes longer because the packets will have a greater number of hops.
2. If the SCFQ scheduling algorithm will be used in establishing IndoEdu-Grid network, then the most appropriate topology is the topology which makes the data packets with the same priorities have little chances to meet each other in a single router or link. In this simulation, topology that meets these conditions is the topology on the Region Based scenario. Topology on the island based and province based scenario does not meet these conditions, so are not appropriate topologies for SCFQ scheduling algorithm.
3. SCFQ scheduling algorithm tends to make packet lifetime in routers with crowded traffic becomes shorter. Packet lifetime shows the difference between the enqueuing and dequeuing time of packets. This is because there are packet priority settings where the packets with higher priority will be served first, so the overall packet lifetime will be reduced.

Based on simulation results in our work, we recommend that the network topology in the Region Based scenario with the implementation of SCFQ scheduling algorithm is used as a reference for establishing IndoEdu-Grid. The Region Based scenario with the implementation of SCFQ scheduling algorithm has the highest level of

effectiveness in terms of job packets propagation, although the achieved level effectiveness is very small (less than 1%). This is because this research used the size and number of Gridlets that are not too large (5000 MI, 3000 MI, and 1000 MI), so the savings or effectiveness only slightly visible.

## References

[1] Berman, F., Fox, G., & Hey, Anthony J.G., 2003. Grid Computing: Making the Global Infrastructure a Reality. West Sussex: John Wiley & Sons.

[2] Buyya, R. & Murshed, M.,2002. Gridsim: A toolkit for the modeling and simulation of distributed management and scheduling for Grid computing. The Journal of Concurrency and Computation: Practice and Experience, 14, 13-15.

[3] Buyya, R. & Murshed, M.,2000. Using the GridSim Toolkit for Enabling Grid Computing Education, http://www.buyya.com/papers/gridsimedu.pdf

[4] Nugroho, I.B. & Suhartanto, Heru, 2010. Design and Simulation of Indonesian Education Grid Topology using Gridsim Toolkit. Asian Journal of Information Technology, Vol 9, Issue 5.

[5] Pustekkom, 2009. Jardiknas – Indonesian ICT Network, http://jardiknas.depdiknas.go.id/index.php/tentang-kami

[6] Sulistio, A ,Poduval, Gokul, Buyya, Rajkumar, and Chen-Kong Tham, 2005., Constructing A Grid Simulation with Differentiated Network Service Using GridSim, In 6th International Conference on Internet Computing (ICOMP 2005), Las Vegas, NV, http://www.gridbus.org/papers/gridsim_net.pdf., also at http://www.buyya.com/gridsim/doc/slides/GridSim_Net_ICOMP2005.ppt

**Heru Suhartanto** is a Professor in Faculty of Computer Science, Universitas Indonesia (Fasilkom UI). He has been with Fasilkom UI since 1986. Previously he held some positions such as Post doctoral fellow at Advanced Computational Modelling Centre, the University of Queensland, Australia in 1998 – 2000; two periods vice Dean for General Affair at Fasilkom UI since 2000. He graduated from undergraduate study at Department of Mathematics, UI in 1986. He holds Master of Science, from Department of Computer Science, The University of Toronto, Canada since 1990. He also holds Ph.D in Parallel Computing from Department of Mathematics, The University of Queensland since 1998. His main research interests are Numerical, Parallel, Cloud and Grid computing. He is also a member of reviewer of several referred international journal such as journal of Computational and Applied Mathematics, International Journal of Computer Mathematics, and Journal of Universal Computer Science. Furthermore, he has supervised some Master and PhD students; he has won some research grants; holds several software copyrights; published a number of books in Indonesian and international papers in proceeding and journal.

**Ivo B. Nurgoho** holds Bsc from Faculty of Computer Sciences University of Indonesia since 2010. He is currently a system integration analyst at Accenture Indonesia and working with Indosat in CRM Project. His main research interest is parallel computing. Together with Heru Suhartanto, he has published an international paper in Asian Journal of Information Technology related with design and simulation of Indo-Edu Grid using Gridsim at 2010.

**Anisa Herdiani** holds B.Sc. and M.Sci from School of Electrical Engineering and Informatics, Bandung Institute of Technology. She is currently an academic and research staff of Faculty of Information Technology – Universitas YARSI. Her research interests are learning technology, consumer health informatics, and knowledge based systems.

# An Enhanced Backoff Method used Between Mobiles Moving in Industrial 802.11

**Walid Fahs[1], Hassan Kabalan [1,3], Jamal Haydar [1], Abbas Hijazi [2], Mourad Gueroui [3]**

**[1] Faculty of Engineering, Islamic University of Lebanon**
**Khaldeh Main Street, Lebanon**

**[2] Faculty of Sciences, Lebanese University**
**Hadath, Beirut, Lebanon**

**[3] Laboratoire d'informatique PRiSM, Versailles St-Quentin-en-Yvelines University**
**45, avenue des États-Unis , 78035 Versailles Cedex, France**

**Abstract**

The purpose of this study is to discuss the exchanges between mobiles moving in an industrial environment. To reach this study, a simulation approach has been chosen to be used. The selection of relevant propagation model for the selected industrial site is based on several measurements. By adjusting the parameters of the various models, we decided to select the model recommended by the ITU under reference Pr1238 to use for our industrial indoor domain.

The second step is to minimize the exchange time between mobiles within an 802.11 cell. The optimization of this time was carried out by modifying the binary exponential aspect of the Backoff algorithm in order to reduce the access time of the radio medium.

***Keywords:*** *WLAN, IEEE802.11, CSMA/CA, BEB, Propagation model AP.*

## 1. Introduction

Wireless local area networks (WLANs) are increasingly popular today. WLANs are used for providing network services in places where it is very difficult or too expensive to lay cabling for a wired network. IEEE 802.11, the most popular WLAN standard, specifications are based on the two lowest layers of the OSI model because they integrate both physical and data link components. All 802 networks have both a MAC and a Physical (PHY) component. The MAC is a set of rules to determine how to access the medium and send data (access method in our case), but the details of transmission and reception are left to the PHY such as sensing the wireless channel and determining whether or not it is idle [1] [2]. Wireless networks use radio frequency channels, as their physical medium in a form of electromagnetic radiation, to exchange data.

Standard wireless networks access methods, especially DCF (Distributed Coordination Function), face some problems related to the transmission priority of different nodes after a success or a fail transmission, these problems cause additional delay [11] [12].

In addition, propagation of radio waves between several nodes moving in the same cell obeys complex rules, especially when there are obstacles between the transmitter and the receiver. A wave can follow several paths to arrive at a common point, so that the receiver can receive multiple copies of the same signal at different instants [3]. The propagation models differ from environment to others, such as industries, universities, home, etc.

Our objective in this study is focused on minimizing the delay of exchanging real-time data between mobiles moving in the same 802.11 cell in industrial domain.

A simulation for several propagation models is performed in NS2 to compare their results to that of the measurement results in the real environment [13].

The rest of the paper is organized as follows: in section 2, we have to pick out the appropriate propagation model for the actual environment. The basic access method DCF is presented in section 3. Section 4 gives an explanation of the enhanced DCF method. In section 5 states the results of the simulations and validation of propagation models. Further, section 6 illustrates the simulations of the enhanced industrial Backoff method and the comparison with the basic method. Finally, section 7 highlights our conclusion and outlines future works.

## 2. Propagation Models

In this section, we present some existing propagation models in the literature and their characteristics. Also, we focused on the "path loss" of each model.
The "path loss" or power dispersion is characterized by the suffered attenuation when an electromagnetic wave traverses a distance. This weakness is mainly due to the dispersion of power and the path obstacles of each received signal component (for example buildings, mountains are signal blockers).

### 2.1 Free space model

This model only assumes the direct path between transmitter and receiver. Eq. (1) represents the received power $P_r$.

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (1)$$

With $P_t$ is the transmitted power, $G_t$ and $G_r$ are respectively the gain of the receiver and transmitter antenna, d is the distance between both nodes and L is the system loss coefficient [3].

### 2.2 Two-ray ground reflection model

The two-ray ground reflection model considers both the direct path and the ground reflection path. The received power at distance *d* is represented by Eq. (2).

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L} \quad (2)$$

Where $h_t$ and $h_r$ are respectively the heights of transmit and receive antennas.
The two-ray model does not give a good result for a short distance due to the oscillation caused by the constructive and destructive combination of the two rays [4].

### 2.3 Log-distance path loss model

The log-distance path loss model is a propagation model that takes into account the different obstacles present in multiple transmitter-receiver paths with the same separation. This model predicts the received signal strength between clients and access points [5].
The log-distance path loss model is given by Eq. (3) [6].

$$L_{total} = PL(d_0) + N.\log_{10}(d/d_0) + X_s \quad (3)$$

Note that:
- *Ltotal* is expressed in decibel,
- *PL(d0)* is the path loss at the reference distance, usually taken as (theoretical) free-space loss at 1m,
- *N/10* is the path loss distance exponent,
- *Xs* is a Gaussian random variable with a mean of zero and a standard deviation of σ dB.

### 2.4 ITU-R P.1238-4

"ITU-R P.1238-4" deals with data propagation and prediction methods for indoor radio systems and radio local area networks in the frequency range from 900 MHz to 100 GHz. As the propagation conditions will vary from one site to another, therefore we should refer to any kind of average site.
The attenuation (power dispersion) is given by Eq. (4).

$$L_{total} = 20\log f + N\log_{10} d + Lf(n) - 28dB \quad (4)$$

With:
- *N* represents the coefficient of attenuation,
- *f* is the frequency in *MHz*,
- *d* is the distance in *meter*,
- *Lf* is the attenuation factor,
- *n* is the number of floors,
- and σ is the standard deviation.

Note that *n* and *N* differ by a factor of 10, as well as in "Free space" model *n* is equal to 2 and *N* is equal to 20 [7].

## 3. Access methods

A station must first gain access to a shared radio channel before transmitting a frame. In 802.11 Standard, two forms of medium access are defined: first one which is mandatory called the Distributed Coordination Function (DCF), and the second one is optional known as Point Coordination Function (PCF) which is only usable on infrastructure network configurations [8]. When applying a PCF method, a Point Coordinator (PC) resides at the Access Point of the Basic Service Set (BSS) and controls frame transfers during a Contention Free Period (CFP) to determine which STA has the right to transmit at the current instant by performing the role of the polling master. The PC controls the frame transmissions of the STAs to eliminate contention for a limited period of time [8]. Our work focuses only on DCF since the infrastructure BSS is used in our environment.
Distribution Coordination Function (DCF) is the fundamental access method of the IEEE 802.11 protocol which is also known as *carrier sense multiple access with collision avoidance* (CSMA/CA). DCF allows for the automatic medium sharing between compatible PHYs through the use of CSMA/CA and a random backoff time following a busy medium condition **[**8**].** Before initiating a transmission, the station senses the medium, if it is idle throughout an interval of time equal to Distributed InterFrame Space (DIFS), the transmission will progress; otherwise it defers until the ongoing transmission terminates and the medium maintains to be idle for a period equal to DIFS, than the station generates a random backoff interval to reduce the probability of collision with

other stations that are also transmitting frames *or* those that are trying to access the same channel.

The DCF implements a slotted binary exponential backoff; when a station finds the medium busy or when it fails to transmit data, it selects a random backoff time. This guarantees that the stations seeking for the channel don't transmit simultaneously. The station sets the Backoff timer using Eq. (5).

$$Backoff\ Time = Random() * SlotTime \quad (5)$$

Where *Random()* returns a random integer within the range *[0,CW]*.

The random delay causes stations to wait different periods of time and prevents them from sensing the medium together at exactly the same instant, finding the channel idle, transmitting, and colliding with each other **[**9**]**. *CW* (Contention Window) is the total time that a source station waits before sending frames, and it's constrained between $CW_{min} \leq CW \leq CW_{max}$, $CW_{min}$ is the initial parameter of *CW* at the first transmission. After each unsuccessful attempt to transmit, *CW* increments double its previous value (2* *CW*) till it reaches its maximum value $CW_{max}$, and it resets to $CW_{min}$ after each successful transmission.

The highest probability for a collision to occur is when multiple stations contend to access an idle channel after it was busy with another station. Here, the backoff procedure is required to avoid the collision, and the station that has the lowest backoff time has the priority to access the channel which consequently decreases the delay of transmission time and optimizes the throughput. If the station fails in transmission, it resets the backoff timer to a new random number and counts down again; CW gets larger as frames fail in transmission $CW = CW * 2^n$. Where *n* is the number of collisions [10].

Some frames fail to reach the destination more than once; therefore the source keeps retransmitting them till these frames reach their destination successfully. Errors that occur, due to a collision between frames or to transmitting either data frame or ACK frame, cause delay in time and lower performance in throughput. These effects are caused by the recovery procedures of the effected frames while exchanged between stations. Additionally, increasing the number of mobiles that are contending on the shared channel ends up with lower performance too.

There are two possible cases in the basic DCF access method: First case is when a collision occurs; the priority to access the channel will be given to one of the remaining contending stations based on its smaller backoff time interval, as mentioned before. This causes the collided frame to delay in transmitting and makes it waits more, since its *CW* is doubling-up.

The second case is when a transmission successes; therefore, *CW* for the transmitting station is equal to $CW_{min}$, as long as it is transmitting successfully, which keeps the priority to this station to access the channel, while on the other side a delay on transmission will arise for the other contending stations.

Based on those two cases, we observe that there are some disadvantages of using IEEE 802.11 DCF scheme, because it suffers a fairness problem [11] [12]. That means that the nodes with smaller *CW* have more chances to get access to the medium, and after a successful transmission the priority to access the medium raises. This issue affects the performance of the CSMA/CA protocol and causes a delay in time.

The Mealy graph shown in Fig. 1, represents the mechanism of BEB.

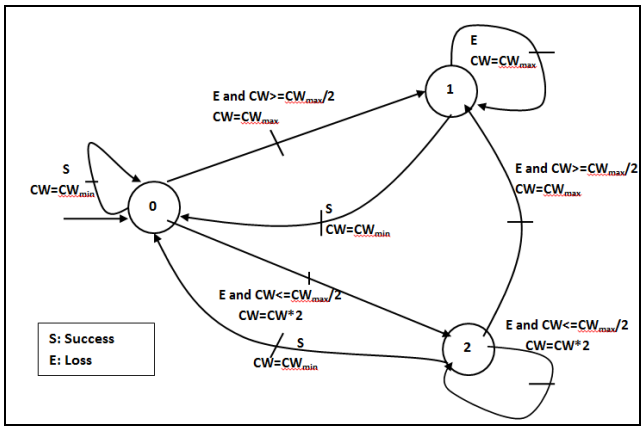



Fig. 1 Mealy graph represents the mechanism of BEB

## 4. Proposed method

This section focuses on an enhanced method of the basic DCF purposed for decreasing the delays that happened in either of the mentioned issues in cases of collision or successful transmission.

This issue is treated by changing the way that *CW* will be set in the two cases:

Firstly when a collision occurs, instead of multiplying *CW* by 2, it is multiplied by *a***,**

Where $0 < a < 2$, thus $CW = CW * a$**.**

Secondly when a frame is successfully transmitted, instead of keeping *CW* of the successful transmitted station equals to $CW_{min}$, we assume that *CW* decreases by *b*, where $0<b<2$, thus $CW = CW - b$.

The following Mealy graph represents the modified BEB method.

The Mealy graph shown in Fig. 2, represents the modified BEB.

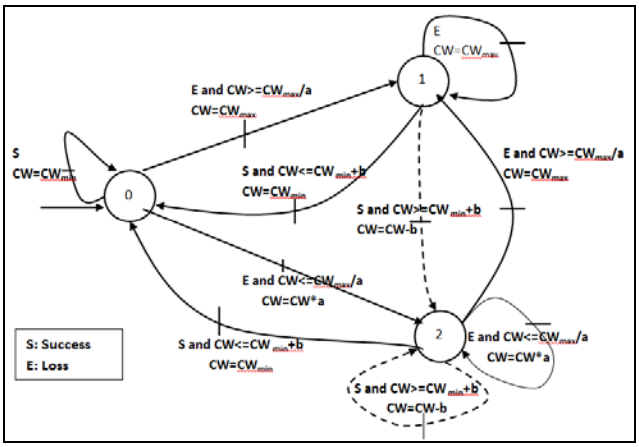



Fig. 2 Mealy graph represents the modified BEB method

## 5. Simulations and validation of propagation model

### 5.1 Selection of the propagation model

In order to minimize the transmission cycle between mobiles, we need to choose the best access method; for that a simulation for several propagation models is performed in a previous work [13], to compare their results to that of the measurement results in the real environment. This comparison shows that the ITU model is the best model that gives propinquity between simulations results and the measurement results.

### 5.2 Implementation of the ITU1238 model

In NS2 [14] [15], the default model used is the "Free space". It expresses the form of the Path Loss Relationship between the received power $P_r$ and the transmitted power $P_t$, this relationship is given by Eq. (1).
To assure that the ITU model matches our work environment, we modify the "Free space" model by replacing Eq. (1) by the equation of the ITU1238 model mentioned in Eq. (4).

### 5.3 Validation of the propagation model

To validate the propagation model (ITU1238), we must compare the simulation results with the power measurements done on the site.
The objective of this simulation is to validate the model implemented in NS2 by comparing the obtained results with $P_r$ measurements on site.

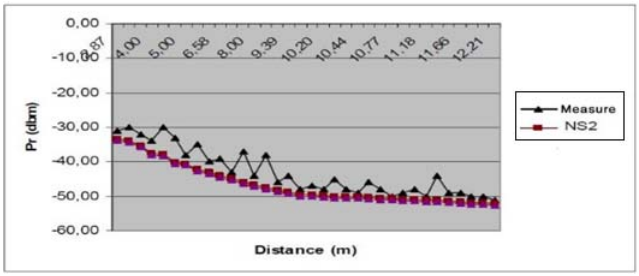



Fig. 3 Results obtained by NS2 and by measurements.

To validate the ITU propagation model, consider the difference between the curve obtained by NS2 and all measurements made on the site. If the difference is approximately equal to zero, thus, we consider that this model is valid and the implementation is realized correctly.
Fig. 3 shows that the average of differences between NS2 results and measurements values is approximately equal to zero.
This propagation model will be used in the next simulations taking into consideration the specific propagation parameters related to the environment.

## 6. Simulation of the enhanced industrial Backoff method and comparison with the basic method

To increase the performance of BEB, we try to change the values of *a* and *b*, where the value of *a* is set to a fixed number of time slots between 1 and 2. For each fixed value of *a*, the value of *b* varies between 0 and 2, in aim to find the lowest time delay (msec) at high packet arrival rate. Fig. 4 shows that the optimum values of delay time (approximately about 50 msec) and packet arrival rate (100) is obtained when $a = 1.4$ and $b = 1.2$

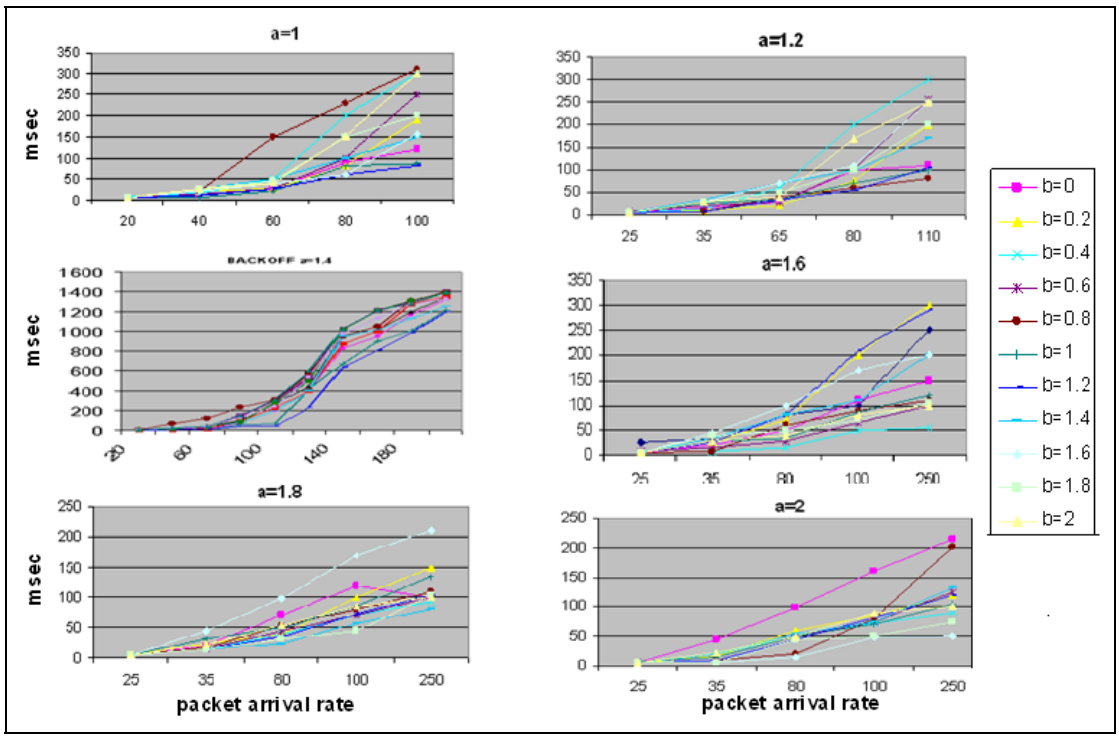



Fig. 4 Results of simulation of modified BEB by varying a and b.

The graph presented in Fig. 5 compares the results obtained by the BEB method and the results obtained by the enhanced method. The transmission time of the enhanced BEB method is less than the one of the BEB method for all values of packet arrival rate.

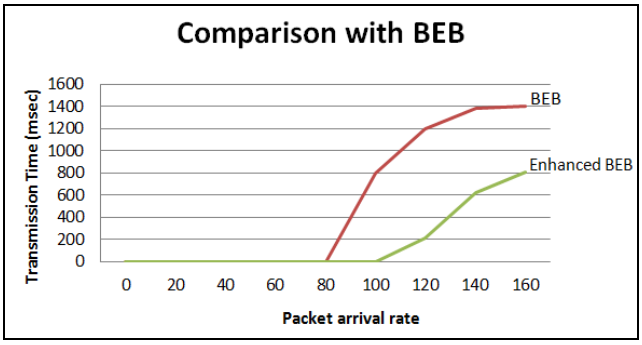



Fig. 5 comparison of the results obtained by the BEB method and the results obtained by the enhanced method

## 7. Conclusion

In this paper, after comparing the simulation results from NS2 for different propagation models to the real environment measurement results, ITU shows the best performance that fits to our industrial environment (indoor environment). In addition, those parameters of the ITU model have been used in NS2 to make the simulation for the standard and the novel access method which performed better time delay. DCF standard model has some disadvantages when it comes to time delay caused by the collisions or by the alternative success transmissions meaning that the mobiles suffer a fairness problem as mentioned. A new method was proposed in order to decrease the time delay where *CW* is multiplied by **a** instead of 2 when a collision occurs and subtract **b** when a transmission successes. The simulation of this new method shows better performance than the basic method DCF particularly when **a=1.4** and **b=1.2** where the time delay was less than 50 msec at 100 packets arrival.

The future work focus on a novel method that is based on the DCF concept, CSMA/CA, but instead of having $CW*2^n$ as a Backoff time, it will have a **Time Priority,** also instead of the mobile waits a DIFS to start sending data, it will wait AIFS time, with ***SIFS*** < ***AIFS*** < ***DIFS***. In this method, a priority is given for each mobile: as its priority increases as the mobile should wait less time. If any mobile (M) wants to send a frame over the network, it should wait a time $t_n$ as it has a priority $P_{n.}$, where n is the priority number.

## References

[1] M. Gast, "802.11 wireless networks: the definitive guide," Publisher: O'Reilly, April 2002.

[2] C. Murthy, B. Manoj, "Ad Hoc Wireless Networks Architectures and Protocols", Publisher: Prentice Hall, Pub Date:2004.
[3] T. S. Rappaport, "Wireless communications, principles and practice", Prentice Hall, 1996.
[4] Friis, H.T., "A Note on a Simple Transmission Formula", 1946.
[5] D.B. Faria, "Modeling Signal Attenuation in IEEE 802.11Wireless LANs - Vol. 1", Technical Report TR-KP06-0118, Kiwi Project, Stanford University, July 2005.
[6] J. S. Seybold, "Introduction to RF Propagation", Published by: John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.
[7] Recommendation ITU-R P.1238-4, "Propagation data and prediction method for the planning of indoor radio communication systems and local area networks in the frequency range of 900 MHz to 100 GHz", 2005.
[8] "Wireless LAN Media Access Control (MAC) and Physical Layer (PHY) Specification", IEEE Standard 802.11, 2007.
[9] J. Geier, "Wireless Networks first-step", Publisher: Cisco Press, 2004
[10] P.S. Kritzinger, H. Msiska, T. Mundangepfupfu, P. Pileggi, A. Symington, "Comparing the results from various performance models of IEEE 802.11g DCF", Computer Networks, vol.54, pp 1672–1682, July 2010.
[11] A. Duda, "Understanding the Performance of 802.11 Networks," Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on, References Cited: pp:1-6, 14 September 2008.
[12] A. Nafaa1, A. Ksentini2, and A. Mehaoua, "SCW: Sliding Contention Window For Efficient Service Differentiation in IEEE 802.11 Networks", In Wireless Communications and Networking Conference, 2005 IEEE, vol. 3, pp: 1626 –1631, March 2005.
[13] F. Walid, "Diffusion d'informations partagées entre mobiles coopérants évoluant sous une même cellule d'un réseau sans fil avec infrastructure ," 2008.
[14] Tutorial for the Network Simulator NS, http://www.isi.edu/nsnam/ns/tutorial/, 2010.
[15] http://www.isi.edu/nsnam/ns/.

# Privacy Preserving RFE-SVM for Distributed Gene Selection

**Fodé Camara[1], Mouhamadou Lamine Samb[1], Samba Ndiaye[1] and Yahya Slimani[2]**

**[1] Department of Mathematics, Cheikh Anta Diop University
Dakar, Senegal**

**[2] University Department of Computer Science, Faculty of Sciences of Tunis
1060 Tunis, Tunisia**

**Abstract**

The support vector machine recursive feature elimination (SVM-RFE) is one of the most effective feature selection methods which has been successfully used in selecting informative genes for cancer classification. This paper extends this well-studied algorithm to the privacy preserving distributed data mining issue. For gene selection over multiple patient data from different sites, we propose a novel RFE-SVM method which aims to learn global informative gene subset to get the highest cancer classification accuracy, with limits on sharing of information. We experiment it using Leukemia bio-medical dataset. The experimental results show that it can provide good capability of privacy preserving and generates a set of attributes that is very similar to the set produced by its centralized counterpart.

***Keywords:*** *Privacy Preserving, Gene selection, Distributed Data Mining, RFE-SVM, Cancer Diagnostic.*

## 1. Introduction

Recently, advances in computing, communications and current hardware technologies have made it possible to collect and store large amounts of data in digital form. For example, high throughput data acquisition technologies have resulted in gigabytes of gene expression data being gathered at steadily increasing rates in biological and bioinformatics sciences. This increasing ability to track and collect large amounts of data has created tremendous opportunities for knowledge-based detecting patterns.

Medical databases are often ideal candidates for large scale, and thus candidates for possibly distributed data mining applications. In fact, data mining over multiple data sources has become an important practical problem with applications in different areas. Due to the sensitive characteristics of personal health records, privacy concern is taken more seriously than other data mining applications. For example, different bioinformatics companies may wish to coordinate themselves in knowing aggregate trends. However, due to privacy concerns, their medical records cannot be brought together. Then, privacy preserving data mining (PPDM) over horizontally partitioned data can be used to achieve this.

This paper applied a privacy preserving gene selection for cancer classification over multiple patient data from different sites. For selecting relevant genes in this case, we propose a novel RFE-SVM algorithm. We experiment it using Leukemia bio-medical dataset. The experimental results show that it can provide good capability of privacy preserving and generates a set of attributes that is very similar to the set produced by the traditional RFE-SVM algorithm.

The remainder of this paper is organized as follows: In section 2, we briefly review the privacy preserving distributed data mining problem. Section 3 provides some background on the RFE-SVM algorithm and the secure multi-party problem. In section 4, we present our privacy preserving RFE-SVM approach. In Section 5, we describe the experiments. In Section 6, we analyze the experiment results. Finally, Section 7 concludes with a discussion of the contributions of our proposal and current research plans.

## 2. Related work

In contrast to the centralized model, the Distributed Data Mining (DDM) model assumes that the data sources are distributed across multiple sites. Algorithms developed within this field address the problem of efficiently getting the data mining results from all the data across these distributed sources. Since this different sources of information are often relating to human subject, many questions concerning their privacy are raised. For example, different superstores with sensitive sales data may wish to coordinate among themselves in knowing aggregate trends

without leaking the trends of their individual stores. As another example, we consider a center for disease control which may want to use data mining to identify trends and patterns in disease outbreaks, such as understanding and predicting the progression of a flu epidemic. Insurance companies have considerable data that would be useful for such a task, but privacy considerations prevent them from releasing the data. An alternative is that each organization performs local operations on its site; this produces intermediate data that can be used to obtain the data mining results, without revealing the private information at each site.

There are many variants of this problem, depending on how the data is distributed, what type of data mining we wish to do, and what constraints are made on shared information.

In all these alternatives, we are in front of two contradictory problems. How to solve the point of conflict between the desire to find a model starting from the union of all these databases, and the right which has an individual to preserve information relating to his privacy?

The main proposal to solve the problem of Privacy Preserving Distributed Data Mining is the Secure Multi-party Computation. A Secure Multi-party Computation (SMC) problem deals with computing any function on any input, in a distributed network where each participant holds one part of the inputs, while ensuring that no more information is revealed to a participant in the computation than its owner input and the output of the function. Secure two party computation was first investigated by Yao [1, 2] and was later generalized to multi-party computation [3]. For example, in a 2-party setting, Alice and Bob may have two inputs $x$ and $y$, and may wish to both compute the function $f(x, y)$ without revealing $x$ or $y$ to each other. This problem can also be generalized across $k$ parties by designing the $k$ arguments function $h(x_1,.., x_k)$.

This approach was introduced into the data mining community for the first time by Lindell and Pinkas in [4]. It allows two different entities to build a decision tree without none of entities being able to know something about the other. Since, many techniques were suggested in the literature. We can divide those techniques in two groups: (i) distributed algorithms over vertically partitioned data; and (ii) distributed algorithms over horizontally partitioned data. There is a horizontally partitioned when the different sites may have different sets of records containing the same attributes, and a vertically partitioned when the different sites may have different attributes of the same sets of records.

The problem of distributed privacy preserving data mining overlaps closely with a field in cryptography for determining secure multi-party computations. A broad overview of the intersection between the fields of cryptography and privacy-preserving data mining may be found in [5]. Clifton et al. [6] give a survey of multi-party computation methods.

## 3. Preliminaries

We start this section with a subsection summarizing the RFE-SVM algorithm. Then, we continue with preliminaries on secure multi-party computation.

### 3.1 The RFE-SVM algorithm

The well-studied RFE-SVM algorithm [7, 8] is a wrapper feature selection method which generates the ranking of features using backward feature elimination. It was originally proposed to perform gene selection for cancer classification [7]. Its basic idea is to eliminate redundant genes and yields better and more compact gene subsets. The features are eliminated according to a criterion related to their support to the discrimination function, and the SVM [8] is re-trained at each step. RFE-SVM is weight-based method, at each step; the coefficients of the weight vector of a linear SVM are used as the feature ranking criterion. The RFE-SVM algorithm [7] can be broken into four steps:

1. Train an SVM on the training set;
2. Order features using the weights of the resulting classifiers;
3. Eliminates features with the smallest weight;
4. Repeat the process with the training set restricted to the remaining features.

### 3.2 The Secure Multiparty Computation problem

Consider a set of parties who do not trust each other, nor the channels by which they communicate. Still, the parties wish to correctly compute some common function of their local inputs, while keeping their local data as private as possible. This, in a nutshell, is the problem we wish to solve, privacy-preserving data mining, is a special case of the secure multiparty computation problem. Before proposing algorithm that preserves privacy, it is important to define the notion of privacy. The framework of secure multiparty computation provides a solid theoretical underpinning for privacy [3]. The key notion is to show that a protocol reveals nothing except the results.

## 4. Proposed Approach

### 4.1 Problem Definition

An inherent tension lies between using medical records for legitimate clinical research and concerns about patient privacy. Consider the example of two different bioinformatics companies. They want to coordinate in knowing relevant genes for cancer classification over the union of their patient data. Due to privacy concerns, their medical records cannot be brought together. In this framework we propose a privacy-preserving distributed RFE-SVM which aims to protect the privacy of patients while maintaining researchers' ability to analyze globally specific genes.

In [9], Fang and al. proposed architecture with a good capability of privacy preserving for decision tree learning. We extend this work to the Feature Selection field. To our knowledge, this work is the first attempt to understand different aspects of using RFE-SVM algorithm over distributed data.

### 4.2 Security tools

Computation on encrypted data does not make sense unless the encryption transformation being used has some homomorphic properties. To define our distributed RFE-SVM algorithm, we use the additive homomorphic encryption and decryption scheme defined in [9].

The encryption scheme is as follows:

- The algorithm uses a large number $r$, such that $r=p\times q$, where $p$ and $q$ are large security prime numbers.
- Given $x$, which is a plaintext message, the encrypted value $y=E_p(x) = x+p \ (mod\ r)$.

The decryption scheme is as follows:

- Given $y$, which is a ciphertext message, we use the security key $p$ to recover plaintext $x=D_p(y) =y \ mod\ p$.

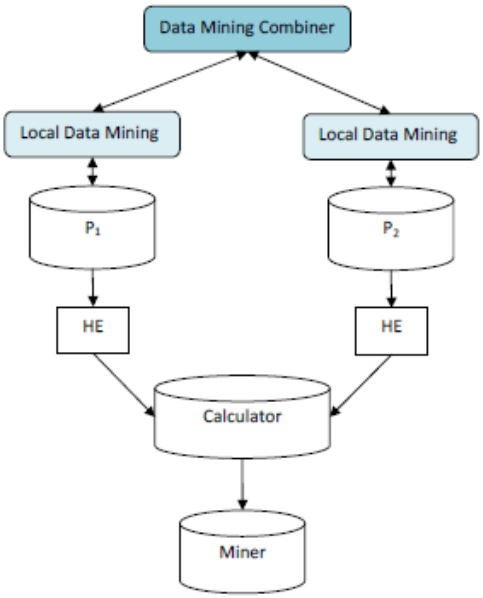



Fig. 1 Secure Multiparty Computation Architecture.

### 4.3 Algorithm

Assume that there are two parties named $P_1$ and $P_2$ which respectively has $m_1$ and $m_2$ sample records, and want jointly selecting the most relevant genes for cancer classification. As described in architecture (c.f. Figure 1), our privacy preserving algorithm is composed of three parts (Algorithm 1, 2, and 3).

## 5. Experimental studies

### 5.1 Experimental Set Up

The experiment was conducted with dual core 2.20 GHz with 4.00 Go of memory on windows platform, and we implemented the distributed algorithm using Java Agent DEvelopment (Jade) framework [10] and the Weka API [11].

**Input:** Local training set and the primary key *pk*
**Output:** Encrypted ranking scores
**Step 1:** Initialize the survived Gene subset $G_S = \{a_{1,..,} a_n\}$, with $a_i$ the attribute at position *i*.
**Step 2:** Repeat until all features are ranked:
(a) Train a SVM on the local training set with genes in $G_S$.
(b) Compute *w*, the weight vector of the resulting local classifier
(c) Compute $c_j$, the ranking scores for genes in $G_S$: $c_j=(w_i)^2$
(d) Send $Enc_{pk}(c_j)$ to the Calculator
(e) Receive the smallest ranking score in $P_1 \cup P_2$ from the Miner, where $P_1$ and $P_2$ are the two sites.

Algorithm. 1 Pseudocode of algorithm performed by the two parties

**Input**: All the encrypted ranking scores from $P_1$ and $P_2$
**Output**: The sum of the encrypted ranking scores
**Step 1:** Receive all $T_1[j]=Enc_{pk}(c_j)$ from $P_1$ and all $T_2[j]=Enc_{pk}(c_j)$ from $P_2$
**Step 2:** Compute $T[j]= T_1[j]+ T_2[j]$
**Step 3:** Send the array *T* to the *Miner*

Algorithm. 2 Pseudocode of algorithm performed by the Calculator

**Input:** An array *T*, which contains the sums of the encrypted ranking score
**Output:** The gene with the smallest ranking score
**Step1:** Receive *T* from Calculator
**Step2:** Decrypt each *T[i ]* using the security key *pk*.
**Step3:** Find the gene with the smallest ranking score and send it to $P_1$ and $P_2$.

Algorithm. 3 Pseudocode of algorithm performed by the Miner

### 5.2 Dataset

To demonstrate real practicality of our approach, we ran experiments on Leukemia bio-medical dataset. The Leukemia dataset consists of samples from patients with either acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML). It initially contains expression levels of 7129 genes taken over 72 samples. Training dataset (Train), given to select genes and adjust the weights of the classifiers, consists of 38 samples (27 ALL and 11 AML). Also an independent test set is provided to estimate the performance of the classifiers. It contains 34 samples (20 ALL and 14 AML).

In our experiments, we only use the 1000 informative genes. To select these, we use '*InfoGainAttribute-Eval*' algorithm and Search Method "*Ranker*" of Weka API[11].

The same testing dataset is used in the two sites, and we partition the training dataset into two parts:

- In site $P_1$: The local training dataset (Train1) contains 19 samples (14 ALL and 5 AML).
- In site $P_2$: The local training dataset (Train2) also contains 19 samples (13 ALL and 6 AML).

This horizontally distributed training dataset (Train) has the following propriety: Train1 $\cup$ Train2=Train where $\cup$ denoted the set union operation.

## 6. Results and discussion

From the horizontally distributed data described above, we conduct experiment to evaluate the performance of our method. Then we compared its performance with the traditional RFE-SVM which is often considered as one of

the best gene selection algorithms in the literature [7]. The comparison results are shown in Figures 2, 3 and Tables 1, 2, 3. The Figure 2 and the Table 1 display the classification accuracy, varying with the number of genes, between traditional RFE-SVM (non-privacy preserving approach) and our approach (privacy preserving approach in distributed feature selection). Table 2 shows the selected gene subsets and their classification accuracy obtaining by using the traditional RFE-SVM approach and our approach, respectively, from which we can see that although the subset returned by RFE-SVM is smaller than that returned by our privacy preserving algorithm, the performances of RFE-SVM and PPD RFE-SVM classifiers are the same.

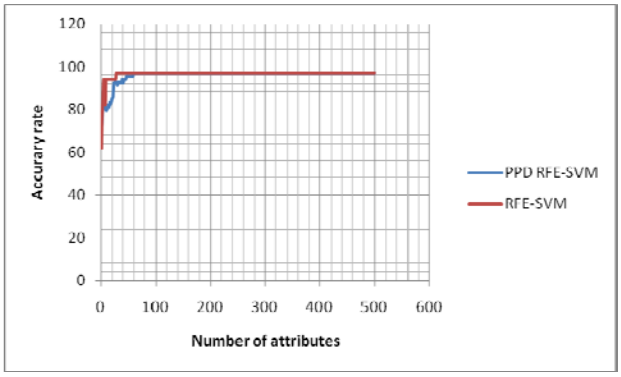



Fig. 2 The classification accuracy varying with the number of genes.

It is obvious that our privacy preserving process increases the complexity computational. But computation cost does not constitute really a problem because there are several parallel architectures.

Table 1: Accuracy rate for different values of N

| ***N*** | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 1000 |
|---|---|---|---|---|---|---|---|---|
| ***RFE-SVM*** | 0.88 | 0.94 | 0.94 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| ***PPD RFE-SVM*** | 0.81 | 0.81 | 0.85 | 0.91 | 0.93 | 0.95 | 0.97 | 0.97 |

Table 2: Performance comparisons between RFE-SVM and PP RFE-SVM

| ***Algorithm*** | ***The smallest number of genes*** | ***Accuracy rate*** |
|---|---|---|
| RFE-SVM | 28 | 0.97 |
| PPD RFE-SVM | 60 | 0.97 |

The Figure 4 and 5 highlight the effectiveness of our distributed gene selection method. Figure 4 shows how informative genes generated by our privacy preserving are similar to those returned by the traditional RFE-SVM. At least 60% of genes are the same. We also see that the top 5 relevant gene subsets obtained are the same. In Figure 5, we display the 10 informative genes selected by the traditional RFE-SVM and our privacy preserving RFE-SVM. Note that the optimal *k* relevant gene subset is not unique, because of the combinatorial nature of the gene selection problem.

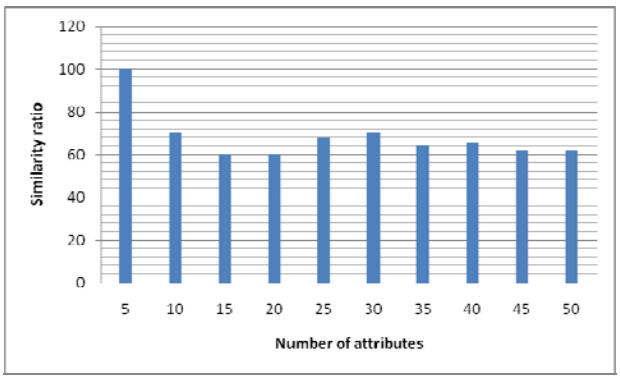



Fig. 3 Similarity between subsets returned by traditional RFE-SVM and PPD RFE-SVM

Table 3: The 10 gene subsets.

| Algorithm | Selected gene subset | Accuracy |
|---|---|---|
| RFE-SVM | {4847, 1834[a], 1779[a], 6539[a], 5772[a], 461[a], 5039[a], 3847[a], 2001, 6184} | 0.94 |
| PPD RFE-SVM | {1834[a], 1779[a], 1745, 6539[a], 5772[a], 4499, 461[a], 5039[a], 3847[a], 1394} | 0.81 |
| **[a] The genes present in the two subsets** | | |

## 7. Conclusions and Future Work

In this work, we proposed a privacy-preserving RFE-SVM for distributed gene selection. It aims to select, over multiple data sources, the smallest informative gene subset to get the highest cancer classification accuracy. To our knowledge, this paper is the first attempt to understand different aspects of using RFE-SVM algorithm over

distributed data. Our experimental results show that our approach has a good capability of privacy preserving, accuracy and efficiency.

In the future, we plan to run experiments on others bio-medical datasets. We also plan to investigate the possibility of reducing the size of smallest informative gene subset while keeping the classification accuracy.

## References

[1] A.C.C. Yao, Protocols for secure computations, Proc. of the 23rd Annual IEEE Symposium on Foundations of Computer Science, Chicago, Illinois, November 1982, pp. 160-164.

[2] A.C.C. Yao, How to generate and exchange secrets, Proc. of the 27th Symposium on Foundations of Computer Science (FOCS), Toronto, Canada, October 1986, pp. 162-167.

[3] W. Du, M.J. Atallah, Secure multi-party computation problems and their applications: a review and open problems, New Security Paradigms Workshop, Cloudcroft, New Mexico, September 2001, pp. 11-20.

[4] Y., Lindell, B. Pinkas, Privacy Preserving Data Mining, In Advances in Cryptology - CRYPTO 2000, pp. 36-54. Springer-Verlag, August 20-24 2000.

[5] B. Pinkas, Cryptographic Techniques for Privacy Preserving Data Mining, ACM SIGKDD Explorations 4(2) (2002).

[6] C. Clifton, M. Kantarcioglu, Tools for privacy preserving distributed data mining, SIGKDD Explorations 4(2), 28-34 (2003).

[7] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Transactions on Knowledge and Data Engineering (TDKE), 17(4), pp. 491-502, 2005.

[8] Y. Tang, Y. Q. Zhang, Z. Huang, Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4(3), pp. 365-381, 2007.

[9] W. Fang, B. Yang, D. Song, Preserving Private Knowledge In Decision Tree Learning, Journal of computers, 5(5), May 2010.

[10] JADE, Java Agent Development framework, http://jade.cselt.it/

[11] H. W. Ian and F. Eibe. Data Mining Pratical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, October 1999.

# A Two Phase Approach for Process Mining in Incomplete and noisy Logs

**Roya ZarehFarkhady[1], Seyyed Hasan Aali[2]**
**[1]Department of Computer Science, Bostanabad Branch, Islamic Azad University, Bostanabad, Iran**

**[2] Department of Computer science, Bostanabad Branch, Islamic Azad University, Bostanabad, Iran**

**Abstract**

***The purpose of process mining is extracting knowledge from even logs recorded in executive information systems.*** ***In many real life logs, too many log instances are needed for the mining approach to work properly. In existence papers about process mining just complete or parallelism tasks with large logs were discussed but in this paper noisy and incomplete logs similarly tested. Therefore, another definitions, metrics and algorithms are required to mine event logs with not enough instances. In this paper, a probabilistic approach is proposed to mine event logs when the number of instances is limited In comparison with many existing approaches, based are the results of our experiments, the proposed approach is very robust in mining process logs with high degrees of parallelism, incompleteness and noise.***

***Keywords:*** *process mining, incomplete, noisy, log, parallelism,*

## 1. Introduction

Information systems are widely used to support the execution of business processes. These information systems typically support logging capabilities that register what has been executed in the organization. These produced logs, called event logs, usually contain data about how people and procedures in an organization work [1]. Process mining techniques are used to discover useful information from the event log, the extracted information can be used to deploy new systems that support the execution of business processes, to find out how to analyze and improve the already enacted process. In systems with no explicit model of the underlying process, process mining techniques can be used to discover the process model and even if an existing process model is expected to be executed these techniques help to find out what is really happening in the organization which may differ considerable form what is assumed to happen [2].

Four different perspectives could be distinguished in process mining; control-flow perspective, the organization perspective, the information perspective, and the application Perspective [3]. However the dominant perspective is the so-called control-flow perspective which mines a model of the process, which specifies the ordering relations between tasks in an event log. In order to mine the control-flow perspective, the workflow log should contain a set of records such that: each record is about an event referring to a task in the workflow instance, it is possible to infer the order in which the tasks are executed. Events also may have timestamps representing the time at which the task is done or recorded. these timestamps are used in some process mining works to adding time information to the process model or to improve the quality of the discovered process model [2].

There are some challenging problems in mining an event log [3], [4] three of the most challenging problems are about mining noisy and incomplete logs and also imbalance execution priorities. Real life logs are almost noisy and incomplete. Therefore in order for a mining algorithm to be applicable it should be able to distinguish exceptions from the normal flow of the model, heuristic approaches are proposed to deal with noise and incompleteness, in these approaches to protect the induction process against inferences based on noise, only task-patron-occurrences above a threshold frequency are assumed reliable enough for the induction process [5]. However it seems that in real life situations where some task-patron-occurrences are very rare, too much data is needed for the algorithms to mine the model correctly. A mining approach might be employed to mine workflow logs with different amounts of noise, and when the proposed approaches [5], [6] are employed to mine workflow logs with high amounts of noise they cannot distinguish among the rare cases mentioned above and the noisy log instances.

Since the issues of noise, imbalance and incompleteness are related, the problem discussed above can be considered to be caused by log incompleteness. A log is said to be incomplete if it does not contain sufficient information to derive the process[3]. However different approaches may need different amount of data to extract

the correct model out a single workflow log, therefore an approach is considered to be more robust with respect to noise if it needs fewer log instances to extract the correct model.

In some of the existing works, in order to see the impact of noise, imbalance and incompleteness on the behavior of a mining algorithm, the benchmark event logs are created via simulation, in these works, CPN-Tools [8] are employed to simulate the process execution, and in order to introduce noise into the event log, some noise introducing operations are performed, however the generated event log may not be realistic since there is no time associated with the workflow tasks, However in a real life situation each well-defined task in a workflow model needs some time to be completed, and when different tasks in the model have different distributions for their completion times, in the generated workflow model some task-sequences may be more frequent while some others are very rare. Therefore assigning time with workflow tasks causes the generated workflow log to be more similar to the real life logs.

In this paper a probabilistic approach is presented to deal with the issues of noise, imbalance and incompleteness, in order to remove some of the restrictions in the current approaches we introduce our metrics to decide about the basic dependencies among the tasks. We use the time Petri nets instead of CPN-Tools to generate more realistic workflow logs with different amounts of imbalance and incompleteness; also we use some noise introducing operations similar to the operations used in [5] to generate event logs with different amounts of noise. Our experimental evaluations show that the proposed approach in capable of achieving more accurate results than the many of the present approaches in dealing with noisy and incomplete logs.

## 1. 2. Related Works

The first papers on workflow mining were in the context of software engineering processes. Cook et al. [9], [10], [11], [12], [13] were the first ones to work on process mining, they used three algorithms, Rnet, which was a purely statistical approach, KTail which was a purely algorithmic approach and a Monrovian approach which was a mixture of algorithmic and statistical methods. The Monrovian approach is able to deal with noise and proved to be superior to the other two algorithms. Agrawal et al. [14] were the first ones to apply process mining in a business setting. Their algorithm is able to deal with noise and assumes that each task appears only once in a process instance. However, Herbst et al. proposed an approach [15], [16], [[17] which is able to tackle duplicate tasks. In their works a two-step approach is employed to mine the process models, in the first step the dependencies between the tasks are captured and represented by Stochastic Activity Graph(SAG). In the second step the SAG is converted to a block-structured process model represented by Adonis Definition Language. The SAG models the behavior in the log but does not contain any duplicate tasks. In order to deal with duplicate tasks, the algorithm applies a set of split operations to nodes in the SAG.

Van der Aalst et al. [7], [18], [19], [20], [21], [22], [23], [24] have focused on mining process models. In [13] they developed the α-algorithm and also proved to which class of models their approach is guaranteed to work. In their approach the event log is assumed to be noise-free and complete with respect to a defined notion of log completeness. However, the proposed approach had problems in mining some common constructs in workflow models. Among these constructs are short loops, which are loops of length one and two. For instance, the α-algorithm was proven to mine sound Structured Workflow nets without short loops.

The α-algorithm does not take into account the frequency of a relation. Therefore, the algorithm is very sensitive to noise and even one erroneous example can completely mess up the derivation of a right conclusion about the binary relations. In addition, the definition of completeness given in this approaches is very arbitrary and strong.

Weijters et al. [23] presented a mining algorithm that uses the ideas behind the α-algorithm and the Cook et al.'s approach. This algorithm uses the frequency of binary relations among tasks to infer the basic relations and is able to mine noisy and incomplete logs. The main idea behind the heuristics is that the more often task A follows task B and the less often B follows A, the higher the probability that A is a cause for B. The algorithm is implemented as the Heuristics miner plug-in in the ProM framework tool [22].

A multi-step approach is introduced by Van Dongen et al. [23]. In their approach, the binary relations like the relations used n the α-algorithm are inferred from the log. Based on these relations, a model is built for each individual workflow instance in the workflow log; the model represents the order between the tasks executed in the workflow instance. Then, in the final step, these instance models are aggregated to obtain an overall model for the entire data set. It is very important to perform a suitable aggregation in the final step.

Most of the existing algorithms in workflow mining are not able to find the long distance dependencies among tasks. Due to the local strategy used in some algorithms

only the local constructs can be mined. However, Wen et al. [23], [21] have proposed two extensions of the α-algorithm, one of these extensions, the β-algorithm [18], is based on the assumption that the tasks in the log are non-atomic. The other extension, the $\alpha^{++}$-algorithm [22], uses the non-local information in the log to mime Petri nets with local or non-local non-free choice constructs.

## 3. The Proposed approach

In this section we present our approach in finding the direct successors. There are some important problems a process mining approach should overcome; dealing with noise, imbalance, incompleteness and mining event logs with high degrees of parallelism. In order to deal with this challenges new concepts and metrics need to be introduced. Since knowing the direct successors the Petri net model can be constructed out, therefore determining the direct successors is essential in mining a workflow model, therefore some metrics and concepts are needed to decide about the direct succession relation.

For each task A, #A denotes the overall frequency of task A in the workflow log, and for each two (different or same) tasks A and B, the following information is used: # AB: the frequency of A directly preceded by B. # A..B: the frequency of A directly or indirectly followed by task B but before the next appearance of A. # A..B..A: the frequency of A directly or indirectly followed by task B but before the next appearance of A and then B directly or indirectly followed by task A but before the next appearance of B, and # A..B..B: the frequency of A directly or indirectly followed by task B but before the next appearance of A and then B directly or indirectly followed by task B but before the next appearance of A.

One of this metrics is the mean distance metric discussed in the following.

Some motivating issues are discussed in this section, showing the usefulness of our approach. In [6] a metric is used to indicate how certain we are that the (A, B) truly belongs to the → relation, the notation is A $\Rightarrow_W$ B and is calculated as follows:

$$A \Rightarrow_W B = \left( \frac{\#AB - \#BA}{\#AB + \#BA + 1} \right)$$

The values of $\Rightarrow_W$ between the events are used to determine the true direct successors, if the value of A $\Rightarrow_W$ B is above a certain threshold, say p, then it is induced that B is a direct successor of A. Although for many dependency relations it is unnecessary to use a threshold value and the all-activities-connected heuristic [6] helps we to take the direct successors but there are many cases in which a threshold value seems to be necessary. The threshold value should be selected with respect to the amount of noise and the degree of concurrency and imbalance in the event log. However different parts of the model may have different degrees concurrency and imbalance, therefore a single value of p cannot be chosen suitable or all parts of the model. This might lead to a wrong derivation even in the simple case in Fig. 1 When due to the completion time distributions of the tasks we have:

$$N = \#BD + \#DB,\ \#DB / N = e$$

And e is a small real number, since for large values of N we have:

$$B \Rightarrow_W D \approx \left( \frac{(1-2e)N}{N+1} \right) \approx (1-2e)$$

Therefore we have B$\Rightarrow_W$ D ≈ (1-2e) and whatever the threshold value, p, is, for small values of e we might have (1-2e) > p and it is derived that D is a direct successor of B.

In [5] some other rules are proposed to decide whether the (A, B) belongs to the → relation, in that approach

a subtle approach is presented to mine noisy and incomplete logs. In their approach some metrics are used to determine the direct successors, the metric #A$\rightarrow_W$ B is used to show the strength of the causal relation between tasks A and B and is calculated dividing the #A$\rightarrow_W$ B-causality counter by the minimum overall frequency of task A and B. One major rule is used in this heuristic approach:

If ((#A→W B > N) and (#AB > θ) and ((#BA < θ)) then B is a direct successor of A

The value θ is automatically calculated using the following equation: θ =1+Round (N×#L/#T). Where N is the noise factor and #L is the number of workflow instances in the workflow log, and #T is the number of tasks.

In estimating the strength of the causal relation between the tasks better results achieved when the metric #A$\rightarrow_W$ B is calculated as follows:

$$A \rightarrow_W B = \frac{A \rightarrow_W B - causality\ counter}{\#A.B}$$

Even if B is a direct successor of A, in an incomplete log there may be no instances in which B directly follows A, such a problem may also rise in models with high degree of concurrency, better criterion may be proposed to decide about causal relation when mining an incomplete log.

Much of the problems with the current approaches are caused by log incompleteness. Log incompleteness can be more serious when dealing with high degrees of concurrency. Determining the concurrent tasks in mining an incomplete event log can be a challenging issue.

Some mining algorithms assume all the information in the event log to be correct. However, in most situations this is not the case, the log may contain noise, incorrectly logged information. Like the method used in [5] we

incorporate noise by performing some noise introducing operations, such as; Deleting the head of the event sequence, Deleting the tail of the event sequence, Deleting a part of the body of the event sequences, Interchanging two randomly chosen events, Shifting a randomly chosen event to the right, and Shifting a randomly chosen event to the left.

We assume (A, B) to belong to the direct succession relation if there are enough instances in which B appears shortly after A. When B is a true direct successor of B, it is likely to have enough instances in the log to show this behavior. If there is no task concurrent to A or B then we expect that B appears directly after A and the distance between A and B in the log is expected to be 1. However if there are tasks concurrent to A or B, the tasks may be appeared between A and B, in the log instances, and if all the occurrences of these concurrent tasks are removed from the log in the resulted log, B appears directly after A and the distance between A and B is resulted to be 1.

Consider again the workflow model shown in Fig. 1. Although the topple (B, C) belongs to the direct succession relation, but if task C requires a long completion time, there may be no instance in which B is directly followed by C. Assume there are 100 log instances containing both B and C, in 56 instances D, in 35 instances E, in 9 instances DE appears between the tasks B and C. Also assume that we know:

$$\{(C,H),(D,H),(E,H),(C,I),(D,I),(E,I)\} \subset \|$$

By removing all the concurrent tasks D and E, 100 cases will be resulted in which the distance between B and C is 1. This helps us to decide that C is a direct successor of B. However, we are never sure if two tasks are in parallel, we only have some probabilities about concurrency between tasks. These probabilities are obtained based on the observations. The expected value for the distance between $T_i$ and $T_j$ is denoted by $\delta(T_i, T_j)$ and calculated as follows:

```
δ(Ti , Tj) = 1;
   for each Tk between Ti and Tj
      if DS(Ti, Tk) and DS(Ti , Tj) then δ(Ti ,
Tj) = δ(Ti , Tj) + (1-P(Tk || Tj))
      else if DS(Tk, Tj) and DS(Ti , Tj) then
δ(Ti , Tj) = δ(Ti , Tj) + (1-P(Ti || Tk))
      else δ(Ti , Tj) = δ(Ti , Tj) + (1-(P(Ti ||
Tk)×P(Tk || Tj)))
   if δ(Ti , Tj) > 3 then δ(Ti , Tj) = 3
```

Where DS(Ti, Tk) is calculated as follows:

$N = \#AB + \#BA$

$M = \#ABA + \#BAB$

$F(A \rightarrow B) = (\#AB - \#BA)/(N+1)$

$F(A \| B) = (1 - |F(A \rightarrow B)|) \times N/(N+1)$

$F(A \# B) = 1 - N/(N+1)$

if $(M > 5\theta)$ then $F(A \square B) = 1$ else $F(A \square B) = 0$

if $(F(A \rightarrow B) > F(B \rightarrow A)$ and $F(A \rightarrow B) > F(A \| B)$ and $F(A \rightarrow B) > F(A \#$

$F(A \rightarrow B) > F(A \square B)$ then $DS(A,B) = true$ else $DS(A,B) = false$

Where $F(A \rightarrow B)$, $F(A \| B)$, $F(A \# B)$ and $F(A \square B)$ indicate how certain we are that the topple (A, B) belongs to the $\rightarrow$, $\|$, #, relation respectively. We set DS(A, B) = true if according to this metrics $F(A \rightarrow B)$ has a larger value than the other three metrics.

When in a workflow model the task B is a direct successor of A and in a log instance C is appeared between A and B then it can be inducted there can be three possibilities; C is a successor of A and is parallel to to B, C is a predecessor of B and is parallel to A, C is parallel to both the tasks A and B. the first possibility shows the motivation for the first line in the body of the for-loop. The task A in this model is an AND-SPLIT. Since the tasks B and C are in parallel, in calculating δ(A, B), the distance A and B, for instances containing the substring ACB, the task C should not be removed. Therefore, if we are pretty sure that both the topples (A, B) and (A, C) belong to the $\rightarrow$ relation, then we have:

$$\delta(A, B) = 1 + (1 - P(Tk \| Tj))$$

A similar reasoning about third possibility can illustrate motivation for the third line in the for-loop body, the motivation behind the second line in the body of the for-loop is explained using the part of workflow model in second possibility. Now consider the distance between A and C, in every log instance B resides between A and C, and since P(A || B) and P(C || B) are both small numbers near to 0, δ(A, C) is about 2.0. Since there may be different path between A and B, therefore we need to see if there are enough cases in which we the distance between A and B is less than 1.1, so we use the notation $\overline{\delta}(A,B)$ to denote the mean distance.

In order to calculate the distance between two tasks we need to decide if two tasks are concurrent, however, we don't know for sure whether two tasks are in parallel, we may just have an estimation of the probability of A and B be in parallel, we tried to propose a metric how sure we are that two tasks, say A and B, are concurrent. Our estimation of this probability is as follows:

$$p(A \| B) = \frac{N}{N+1} e^{-((1.25/(0.5-\theta))(x-0.5))^K},\ x = \frac{\#A.B}{N},\ N = \#A.B + \#B.A$$

The tasks A and B are said to be concurrent if they can be executed in any order, in cases where the event log is assumed to be complete and noise free it is very easy to

determine concurrent tasks, in such a situation tasks A and B are determined to be concurrent if (#AB>0 and #BA>0). In noisy situations decisions may be made based on the frequencies and only task-patron-occurrences above a threshold frequency are reliable enough for our induction process, however when mining an incomplete log, deciding based on the frequencies may lead to incorrect results. In our approach, if both the #A..B and #B..A are above a certain threshold we may conclude that A and B are concurrent; however this is an efficient condition and is not necessary as shown before. The above formula is used to indicate how sure we are that two task are truly concurrent: Where N is summation of #A..B and #B..A and x is the fraction of times A appears before B and θ is calculated according to the following:

$$\theta = 1 + Round(\frac{F \times L}{3T})$$

Where F is the noise factor, L is the number of workflow instances containing both the task A and B and T is the number of task types in the workflow log, K in the above formula can be any even number, in our experiments we set K = 20, the plot for K=10 and N = 80 is shown in figure 3. As one can see for $x < \theta$ and $x > 1 - \theta$ the probability of the A and B to be in parallel is estimated to be zero and only for $(\theta < x < 1- \theta)$ the probability has a non-zero value.

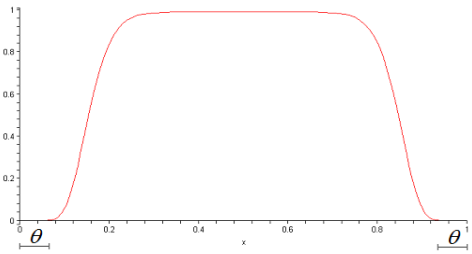


Figure 2. The probability of two tasks being in parallel as a function of #A..B and #B..A

Our estimation of this probability is based on the event log which might be incomplete.

Our algorithm in finding the direct succession relation is based on the simple metric called mean distance among tasks. Using the event log we try to calculate the distance between each two none concurrent tasks in the model, task B is assumed to be a direct successor of A if there are enough instances in which δ(A, B) is about 1.0. We use this rule to decide about the direct successors:

$P(A||B)<0.5$ and $\overline{\delta}(A,B) < 2$ and $F(Support(A \rightarrow B)) > 0.7$

## 4. Experimental Results

In order to evaluate the proposed algorithm, some experiments have been done on five time Petri net models. The models have complexities comparable with the complexity of models used in [6]. The models contain self-loops, length two loops and other kinds of loops. the degrees of parallelism in the models ranges from 2 to 5 execution threads, have not duplicate tasks and no hidden task exists except that all the AND-SPLIT, OR-SPLIT, AND-JOIN and OR-JOIN task supposed to be hidden. number of tasks in this models ranges from 24 to 67 tasks.

All the models have been used with Different degrees of imbalance and different completion time distributions assigned to the tasks, in order to see the impact of noise on our mining algorithm the experiments are done with different amounts of noise in the event log, events log without noise, with 5%, 10%, 20%, 30%, 40% and 50% noise are used to show the behavior of the proposed algorithm. Also the experiments are repeated with different amounts of imbalance in the branch points.

Assigning different completion time distributions to the tasks, different workflow logs are generated and mined.

the completion time assigned to the tasks in our experiments is the summation of two parts, a deterministic random variable randomly chosen in the interval (10, 15) and a normal distribution random variable with $\mu = 25$ and $\sigma^2 = 7$, the results from 100 executions of our mining algorithm with different completion times assigned to the tasks is shown in TABLE 1

.TABLE 1 . RESULT OF ALGORITHM

| Imbalance in branch points: 50% | | Amount of noise | | | | |
|---|---|---|---|---|---|---|
| | | 0% | 5% | 10% | 20% | 30% |
| Number of instances | 10 | 0.919, **0.9** | 0.926, **0.93** | 0.918, **0.92** | 0.914, **0.92** | 0.921, **0.962** |
| | 20 | 0.887, **0.95** | 0.886, **0.951** | 0.893, **0.954** | 0.902, **0.953** | 0.918, **0.95** |
| | 30 | 0.873, **0.960** | 0.882, **0.962** | 0.883, **0.963** | 0.887, **0.94** | 0.912, **0.94** |
| | 50 | 0.872, **0.962** | 0.861, **0.94** | 0.868, **0.962** | 0.884, **0.961** | 0.917, **0.960** |
| | 100 | 0.868, **0.961** | 0.875, **0.959** | 0.876, **0.957** | 0.884, **0.956** | 0.914, **0.953** |
| | 200 | 0.883, **0.954** | 0.892, **0.97** | 0.880, **0.960** | 0.907, **0.952** | 0.928, **0.951** |
| | 300 | 0.916, **0.955** | 0.915, **0.957** | 0.908, **0.960** | 0.919, **0.960** | 0.946, **0.952** |
| | 500 | 0.940, **0.967** | 0.937, **0.961** | 0.931, **0.965** | 0.936, **0.962** | 0.953, **0.959** |
| | 1000 | 0.976, **0.975** | 0.962, **0.970** | 0.973, **0.977** | 0.974, **0.975** | 0.963, **0.969** |
| | 2000 | 0.988, **0.988** | 0.979, **0.979** | 0.976, **0.981** | 0.980, **0.981** | 0.972, **0.974** |

The result of our experiments for different amounts of noise and different amounts of workflow log instances are shown in the TABLE 1, in each experiments the number of correctly determined elements in the Boolean matrix of direct successors relations is divided by the number of all the elements, the results obtained by applying the approach proposed in this paper in compared with the

result of the approach presented in [5], the results of our approach are shown bold. As one As shown in fig 2.

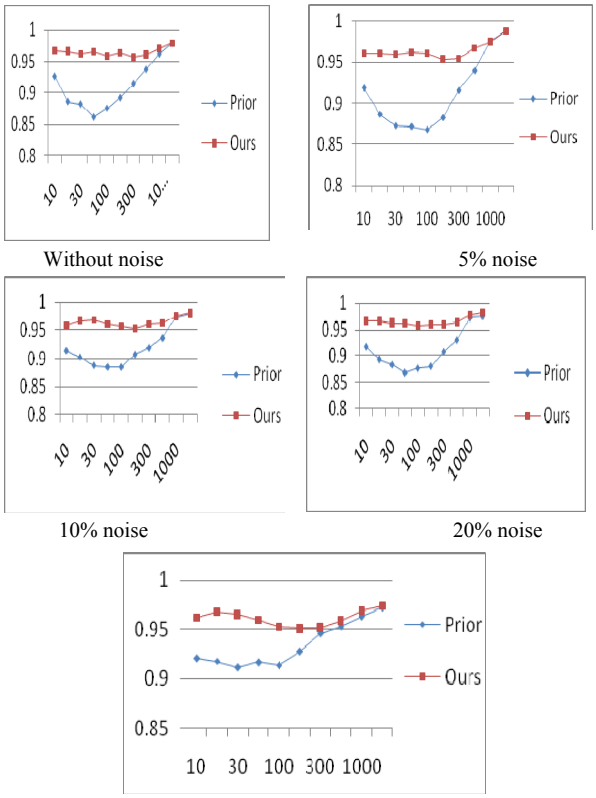



Figure 2. Evaluation results

## 5. Conclusion

The proposed approach uses a two phase algorithm to find the direct successors. New metrics and concepts are introduced to use the existing event log more efficiently.

We have introduced time Petri nets, instead of CP-nets in similar works, to create more realistic logs. We have also shown that our approach is more successful than many of the existing techniques. Our experiments on 250 different workflow models shown that our approach is more robust in mining incomplete and noisy event logs.

## References

[1] W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, A.J.M.M. Weijters, "Workflow mining: a survey of issues and approaches", Data and Knowledge Engineering, 2003.

[2] W.Aalst,. van der, & Weijters, A. "Process mining: A research agenda". Computers in Industry, 2004.

[3] A.J.M.M. Weijters W.M.P van der Aalst, "Process Mining Discovering Workflow Models from Event-Based Data", springer, 2000.

[4] A.J.M.M. Weijters, W.M.P. van der Aalst, and A.K. Alves de Medeiros, "Process Mining with the HeuristicsMiner Algorithm", 2006.

[5] A.K.A. de Medeiros, A.J.M.M. Weijters, and W.M.P. van der Aalst. "Genetic Process Mining: A Basic Approach and its Challenges", Springer, 2006.

[6] W.M.P. van der Aalst and A.J.M.M. Weijters, editors, "Process Mining: Special Issue of Computers in Industry", Elsevier Science Publishers Amsterdam, 2004.

[7] R.Silva, J.Zhang, G. Shanahan, "Probabilistic Workflow Mining", ACM, 2005.

[8] J. Herbst and K. Karagiannis. "Workflow mining with InWoLvE". Computers and Industry, 2004.

[9] R. Bergenthum, J.Desel, R.Lorenz, and S. Mauser, "Process Mining Based on Regions of Languages", Proceedings of the 5th International Conference on Business Process Management, 2007.

[10] B. van Dongen, N. Busi, G. Pinna, and W. van der Aalst. "An iterative algorithm for applying the theory of regions in process mining". Technical Report Beta rapport, 2007.

[11] R.Bergenthum, R. Lorenz, S.Mauser," Towards Applicability of Language Based Synthesis for Process Mining", Proceedings of the 5th International Conference on Business Process Management, 2007.

[12]W. M. P. Aalst, C. W. GÄunther, "Finding Structure in Unstructured Processes: The Case for Process Mining.", In: Proceedings of the 7th International Conference on Application of Concurrency to System Design (ACSD), 2007.

[13] R. Bergenthum, J.Desel, R.Lorenz, and S. Mauser, "Experimental Results on Process Mining Based on Regions of Languages", ICATPN, LNCS, 2008.

[14] J. van der Werf, B. van Dongen, C. Hurkens, and A. Serebrenik, "Process discovery using integer linear programming", ICATPN, LNCS, 2008.

[15] Process mining group eindhoven technical university: Prom-homepage. http://is.tm.tue.nl/ cgunther/dev/prom/.

[16] Z.Huang, A.Kumar, "A Study of Process Mining: Quality and Accuracy Tradeoffs", Proceedings of the 4th Workshop on Business Process Intelligence, 2007.

[17] W. M. P. van der, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. A. de Medeiros, M.Song and H. M. W. Verbeek, "Business process mining: An industrial application", Information Systems, 2007.

[18]L. Măruşter, A. J. M. M. T. Weijters, W. M. P. van der Aalst and A. van den Bosch, "A rule-based approach for process discovery: Dealing with noise and imbalance in process logs", Data Mining and Knowledge Discovery", 2006.

[19]A. K. A. de Medeiros, C. W. Günther, A. J. M. M. Weijters and W. M. P. van der Aalst, "The need for a process mining evaluation framework in research and practice", Proceedings of the 3rd Workshop on Business Process Intelligence, 2007.

[20] B.F. van Dongen and W.M.P. van der Aalst." EMiT: A process mining tool". International Conference on Applications and Theory of Petri Nets, Springer-Verlag, 2004

[21]J. Herbst and D. Karagiannis. "Workow mining with inwolve". Computers in Industry, 2004.

[22] A.K. Alves de Medeiros. "Genetic Process Mining". PhD thesis, Eindhoven, University of Technology, Eindhoven, The Netherlands, 2006.

[23] G. Schimm, "Process miner: a tool for mining process schemes from event-based data", Proceedings of the 8th European Conference on Artificial Intelligence (JELIA), Lecture Notes in Computer Science, 2002.

# A Conventional Authentication in Key Management using Progressive Approach

**Sandosh Sakkarapany[1], Uthayashangar Sundaramourty[2]**
**[1]Assistant Professor**
**Department of Information Technology,**
**Manakula Vinayagar Institute of Technology,**
**Puducherry-605 107,India.**

**[2]Assistant Professor**
**Department of Information Technology,**
**Manakula Vinayagar Institute of Technology,**
**Puducherry-605 107,India.**

**Abstract**

Secure and reliable group communication is an increasingly active research area prompted by the growing popularity of many types of group-oriented applications. The main building block to achieve security in group communication scenarios is management of the secret information that should be known only to group members involved in communication. However, the complete security with the key management for such protocols remains a significant problem. In this paper, a secure group key agreement between the group members is proposed. The main advantage of the proposed system is the continuous improvement in Authentication between the group key members.

***Keywords****: Multicast, Cryptography, TGDH, Group Key management, Continuous Authentication.*

## 1. Introduction

Services such as Stocks publishing, News distribution, Audio / Video conference systems, Collaborative work-flow systems and any software updates, use multicasting information among many people. However, an increasing number of such applications require secure multicast services in order to control the Group Members in a secure way. Group Key Management is a methodology which helps to achieve security in multicasting information over group-oriented communication.

In order to multicast information among a certain group securely, a group key should be shared among all members in the group. Every information packages should be encrypted with the shared group key before they are transmitted. Only the authorized users who have the shared common group key can decrypt the package and retrieve the information. The unauthorized users perhaps received the encrypted packages; they cannot retrieve the information without the group key.

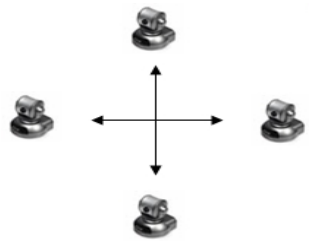


Fig. 1 Many to Many Group Communication

But there are some problems need to be addressed in this way of sharing information. The members in the group may be changed frequently. Any time when a new member joined, a new group key should be generated and distributed to all group members, include the new joined member. This rekeying (generation and distribution of new key) process helps to maintain continuous Authentication to share confidential information among the authorized users.

## 2. Key Management

Multicasting helps the Group Member to send the confidential information to the selected group of recipients in a group. The Key Management can be classified along two primary directions: group key distribution (transport), and group key agreement.

The main difference is that in case of group key distribution there exists a designated participant with extended rights, called group manager or key server that computes the group key on its own and distributes it securely to all other participants, whereas in case of

group key agreement all participants have equal rights and each of them provides own contribution to the computation of the group key.

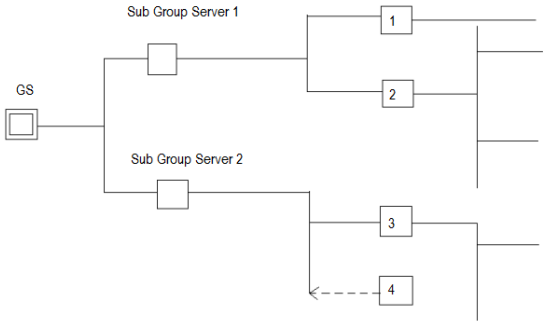



Fig. 2 Key Management

In a Group-Oriented system there is Group Server or Key Distributor, who allows the group members in the group to access data. Initially when any user wants to join a group, the user should request the Group Server that the user is interested in joining the group.

Then the Group Server asks for the one time authentication from the new user which is nothing but Username and Password. After this process the Group Server introduces two keys to the new agent which is called the Group Key and the Individual Key. The Group Key is used to decrypt the confidential information shared among the group members.

All the members in the group have the same Group Key where as the Individual Key is unique and is used to request the Group Server that the particular user in the group is interested in leaving the group. Each Group Server or Key Distributor generates and distributes the Group Key. Then the Group Server encrypts the confidential information with the Group Key and send it to the other member in the group. The group members receive and decrypt the message with the shared Group Key distributed by the Group Server.

Key refreshing is one of the most important security requirements of group key agreement protocols. Because, the session key should be known only to the involved parties.

The following four properties should be guaranteed:

2.1 Group Key Secrecy-It is computationally infeasible for a passive group member who quit the group to discover any group key.

2.2 Forward Secrecy -The new group member who knows a continuous subset of group keys cannot discover any old group key to decrypt the messages.

2.3 Backward Secrecy - A passive group member who knows a continuous subset group keys cannot decrypt the shared information anymore with the key.

2.4 Key Independence -The Individual Key shared between the group server and particular group member is independent.

## 3. TGDH

From all existing group key agreement protocols, the approach called Tree-Based Group Diffie-Hellman (TGDH) protocol suite is chosen for better efficiency. These protocols combine the structure of binary key trees with two-party Diffie-Hellman key exchange protocol to achieve the computation of the secret group key after several rounds. This method is also called an iterative Diffie-Hellman (IDH) key exchange.

This solution is secure, surprisingly simple and also very efficient, compared to other existing group key agreement protocols. TGDH protocol suite is most applicable in dynamic groups, where number of participants changes during the communication period.A group key agreement scheme needs to provide key adjustment protocols stemming from membership changes. TGDH suite includes protocols supporting the following operations:

3.1 Leaving the group -If any member leaves the group rekeying (generation and distribution of new Group Key will be done by the Group Key Server. When a user leaves the group, the Group Key Server generates and distributes the new Group Key.


Fig. 3 Leaving the TGDH tree

3.2 Joining the Group-When a new user enters the group the Group Key Server distributes the new Group Key to the entire users including the new one.

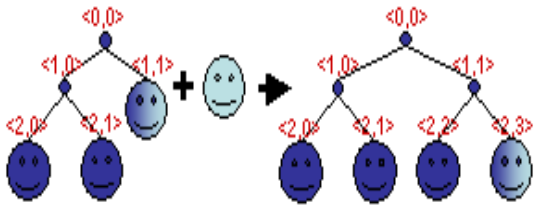



Fig. 4 Joining the TGDH tree

Initially the Group Server is created, then the new group members request the Group Server to enter into the group, then the one time authentication is to be done for each new user successfully. Once the process is done the TGDH Protocol suite forms the group in a binary tree structure for managing the group member in an efficient manner.

## 4. Skipjack

Skipjack is the cryptographic algorithm which encrypts and decrypts data in 64-bit blocks, using an 80-bit key. It takes a 64-bit block of plaintext as input and outputs a 64-bit block of ciphertext. Skipjack has 32 rounds, meaning the main algorithm is repeated 32 times to produce the ciphertext. It has been found that the number of rounds is exponentially proportional to the amount of time required to find a key using a brute-force attack. So as the number of rounds increases, the security of the algorithm increases exponentially.

This cryptographic algorithm used for encrypting and decrypting the confidential information brings more security on the information shared among the group member. Now the information is more secure and the users who access the information are legitimate is to be addressed.

Even though multicasting information over the group oriented environment with the TGDH key management protocol suite, there are still some problems need to be addressed.

Multicast protocol would be subject to attack by an **active intruder** compared to a unicast protocol. There are inherently more opportunities for interception of traffic, would typically make it easier for an intruder to pose as another legitimate principal.

## 5. Proposal

The important problem need to be addressed here is Authentication. Because multicast protocols are easily attacked by active intruders and interception is possible. In order to prevent intruders and maintaining security, a step by step continuous improvement in authentication is required.A multicast protocol is easily subject to attack by an intruder compared to unicast protocol and there are inherently more opportunities for interception.

In the Group Key Server asks for one time authentication (Username and Password) then it generates a new Group Key and Individual Key to the users in the group. After the one time authentication the Group Key Server does not worry about whether the confidential information reaches the authorized destination or user.

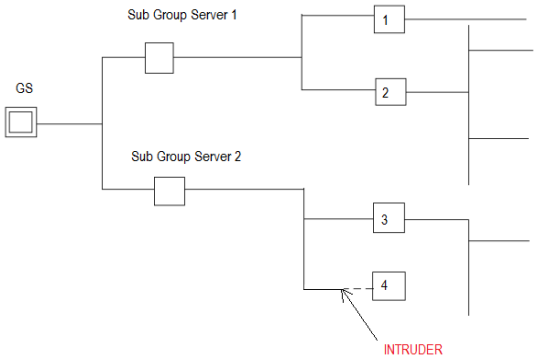



Fig. 5 Intruder

Hence any intruder may receive the encrypted messages from the Group Key Server. Since the intruder do not have the group key he cannot decrypt the encrypted information shared among the group members. But, if any of the users leaves or joins the group, rekeying will be done by the Group Key Server and the intruder will get the new group key then the intruder starts decrypting the confidential information which he has not been authorized to read.

To avoid intruders into the Group, accessing the confidential information, frequent authentication verification is to be done. It is the role of the Group Server keep tracking the information send by the Group Server reaches only the legitimate user. A separate process of cryptography is used to confirm that the users in the Group are original or legitimate user.

### 5.1 Conventional Authentication

Once the one time authentication (requesting the Username and Password) is done the Group Server generates and distributes new Group Key to all the members of the Group and an individual key for the new user. The Group Key is used to decrypt the confidential encrypted messages by all the members and the

individual key is used to inform the Group Server when the member is interested in leaving the Group.
In the proposal, it has been planned that a tiny bit of plain text other than encrypted confidential Group message is sent to all the members in the Group at regular intervals of time. Each member in the Group receives the tiny bit plain text and encrypt it with the unique Individual Key and sends the result of encryption (ie) cipher text back to the Group Server. Then the Group Server also calculates cipher text for that plain text separately in the Server side and compares the resultant value with the received cipher text from each member.

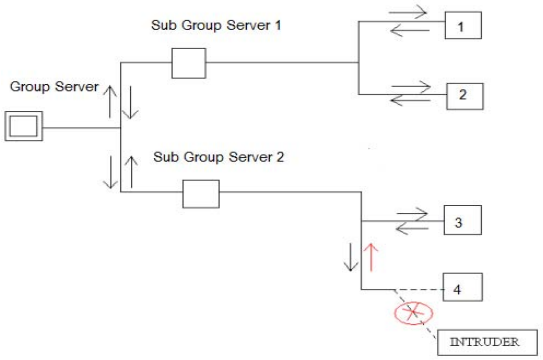



Fig. 6 Conventional Authentication

Any intruder who enters the communication link between the member and the Group Server may get the group key if any one of the member joins or leaves the Group, but the intruder is not possible to get the unique individual key.

Only legitimate group member who is having the unique Individual Key can provide the expected results of cipher text to the server. After receiving the result of encryption from each member the server compares the result.

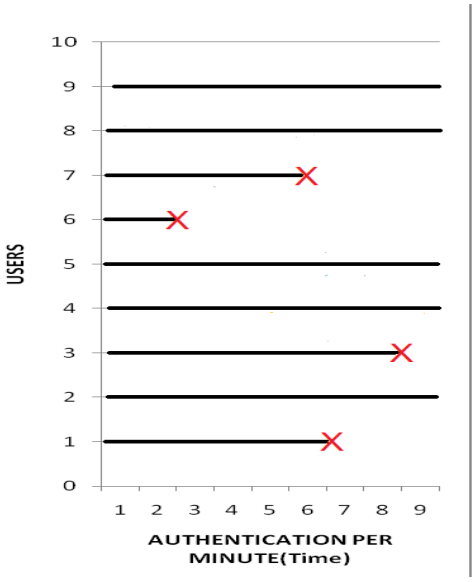



Fig. 7 Step by Step improvement

If match is found with the encrypted result of particular user, the communication continues.

If the result received by the server is identified as wrong one, the server comes to know some unauthorized person is trying to access the information then it stops the communications with the particular group member.

## 6. Conclusion

Sharing confidential information among the Group Members in a secure way is more critical. After the one-time authentication, the new Group Member enters the group and starts sharing the confidential information by Encrypting and Decrypting the messages with the help of the Group Key provided. Skipjack- an efficient Encryption Algorithm helps to maintain the information more secure among the Group Members. During the process of sharing of the encrypted information among the members, it is not assured that the information reaches only the legitimate Group Members. After the one-time Authentication (Username and Password) it is possible for the Intruders to enter the group and receives the confidential information but, the Intruder cannot decrypt the encrypted information without the group key.

In Group Key Management, if any member joins or leaves the group rekeying (generation and distribution of new group key) is done. Hence the Intruder easily receives the Group Key and decrypts the confidential information. Once the new member joins the group, the Group Members shares the information without worrying

about whether the confidential message reaches the legitimate Group Members or not.

With the proposed system, Conventional Authentication is introduced and maintained till the Group Member quits the group. If any Intruder has entered the group, the intruder cannot perform the Authentication Verification process with the Group Server successfully. If the Authentication Verification Process fails, the communication between the Group Server with the intruder is terminated. Hence, the confidential information among the Group Members becomes more secure through conventional Authentication.

## 7. References

[1] Feng Zhu, Wei Zhu, Matt W. Mutka, "Private and Secure Service Discovery via Progressive and Probabilistic Exposure" VOL. 18, NO. 11, IEEE November 2008.

[2] Ling Cheung, Joseph A. Cooley, Roger Khazan, Calvin Newport, "Collusion-Resistant Group Key Management Using Attribute-Based Encryption" MIT Lincoln Laboratory, March 22, 2007.

[3] W. Yu, Y. Sun, and K. J. R. Liu, "Optimizing Rekeying Cost for Contributory Group Key Agreement Schemes," IEEE Transactions On Dependable And Secure Computing, vol. 4, no. 3, pp.228-242, 2007.

[4] F. Zhu, M. Mutka, and L. Ni, "A Private, Secure and User-CentricInformation Exposure Model for Service Discovery Protocols," Mobile Computing, vol. 5, pp. 418-429, IEEE 2006.

[5] F. Zhu, M. Mutka, and L. Ni, "Service Discovery in Pervasive Computing Environments," Pervasive Computing, vol. 4, pp. 81-90, IEEE Oct 2005.

[6] Rahul.S and Hansdah, RC , "An Efficient Distributed Group Key Management Algorithm" Tenth International Conference on Parallel and Distributed Systems, 2004. ICPADS 2004, 7-9 July, California, 230 -237.

**S.Sandosh** received the B.E degree in Computer Science and Engineering from Anna University, India in 2007 and M.Tech degree in Information Security from Pondicherry University, India in 2010. He is working in Manakula Vinayagar Institute of Technology,Puducherry, India as a Assistant Professor in Information Technology Department.

**S.Uthayashangar** received the B.Tech degree in Information Technology from Pondicherry University, India in 2007 and M.Tech degree in Information Security from Pondicherry University,India in 2010. He is working in Manakula Vinayagar Institute of Technology,Puducherry,India as a Assistant Professor in Information Technology Department.

# Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images

**Manavalan Radhakrishnan[1] and Thangavel Kuttiannan[2]**

**[1] Department of Computer Science and Applications, KSR College of Arts and Science**
**Tiruchengode, Namakkal, Tamilnadu, India**

**[2] Department of Computer Science, Periyar University**
**Salem-11, Tamilnadu, India**

**Abstract**

Diagnosing Prostate cancer is a challenging task for Urologists, Radiologists, and Oncologists. Ultrasound imaging is one of the hopeful techniques used for early detection of prostate cancer. The Region of interest (ROI) is identified by different methods after preprocessing. In this paper, DBSCAN clustering with morphological operators is used to extort the prostate region. The evaluation of texture features is important for several image processing applications. The performance of the features extracted from the various texture methods such as histogram, Gray Level Cooccurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GRLM), are analyzed separately. In this paper, it is proposed to combine histogram, GLRLM and GLCM in order to study the performance. The Support Vector Machine (SVM) is adopted to classify the extracted features into benign or malignant. The performance of texture methods are evaluated using various statistical parameters such as sensitivity, specificity and accuracy. The comparative analysis has been performed over 5500 digitized TRUS images of prostate.

***Keywords:*** *GLCM, Histogram, GLRLM, DBSCAN, SVM.*

## 1. Introduction

Digital image plays a vital role in the early detection of cancers, such as prostate cancer, breast cancer, lung cancer, cervical cancer and blood cancer [1]. Prostate cancer is one of the common diagnosed malignancies in middle-aged and elder men, and the survival rate of the patients can only be enhanced by detection in the early stage of cancer [2, 3, 4]. Prostate cancer is now most frequently diagnosed male malignancy with one in every 11 men. It is the second position in cancer–related cause of death only for male population. Ultrasound imaging method is suitable to diagnosis and prognosis. An accurate detection of Region of Interest (RoI) in ultrasound image is crucial, since the result of reflection, refraction and deflection of ultrasound waves from different types of tissues with different acoustic impedance. Usually, the contrast in ultrasound image is very low and boundary between region of interest and background are more uncertain Images are prone to different types of noises. Noise is considered to be any measurement that is not part of the phenomena of interest. And also speckle noise and weak edges make the image difficult to identify the prostate region in the ultrasound image [7]. So, the analysis of ultrasound image is more challenging one. Generally there are two common use of ultrasound medical imaging: first one is to guide the oncologist in the biopsy procedure and second is in the establishing the volume of the prostate. It has been used in diagnosing for more than 50 years. The statistical textural features can be extracted from the region of interest to classify the proteomic into benign or malign. This paper emphasis textural features extraction from Histogram, GLCM, GLRM and their combinations. And Support Vector Machine (SVM) is adopted for classification. The following section presents an overview of the proposed Computer Aided Diagnosis (CAD) system.

### 1.1 Overview of the CAD System

The CAD system consists of five stages such as acquisition of TRUS image of prostate, preprocessing, segmentation, feature extraction and classification. It is developed for automatic detection of prostate tumor in Transrectal Ultrasound (TRUS) image. The overview of the CAD system is depicted in figure 1. It can provide the valuable viewpoint and accuracy of earlier prostate cancer detection.

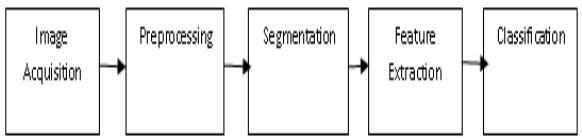



Figure 1: CAD System

The detailed description of each stage is presented in the subsequent sections. The rest of the paper is organized as

follows: section 2 deals about image acquisition and preprocessing of original TRUS medical image of prostate. The DBSACN clustering with morphological operators for locating the Region of Interest (ROI) from TRUS prostate medical image is described in section 3. The feature extraction through Histogram, GLCM, and GLRLM are illustrated in section 4. The SVM classifier is discussed in section 5. The experimental analysis and discussion are presented in the section 6. Section 7 concludes this work with directions for further research.

## 2. Image Acquisitions and Preprocessing

Ultrasound imaging is a widely used technology for prostate cancer diagnosis and prognosis among the different medical image modalities. Image acquisition processes required less than 30 minutes for each case. The TRUS images of the prostate gland are acquired while the TRUS probe is supported and manipulated by the TRUS Robot using a joystick located next to the console, without the need for a dedicated assistant. The entire prostate is scanned by rotating the TRUS probe about its axis, minimizing prostate displacement and deformation. The probe depth can be manually adjusted by a surgeon. Accurate recording of images and corresponding TRUS frame coordinates were obtained. The gathered information is used in offline for the ultrasound image of the prostate gland. Image acquisition is done from ultrasound device via frame grabber. Image of the prostate can be generated using a series of TRUS images.

The ultrasound images are very difficult to process and analysis, because of poor image contrast, speckle noise and missing or diffuse boundaries in the Transrectal Ultrasound (TRUS) images. The data dropout noise is generally referred to as speckle noise [25, 26]. It is, in fact, caused by errors in data transmission. The corrupted pixels are either set to the maximum value, which is something like a snow in image or have single bits flipped over. This kind of noise affects the ultrasound images [8]. To interpret the TRUS images, preprocessing is necessary to improve the quality of image and make the subsequent phases as an easier and reliable one. The M3-Filter is used to generate a despeckled image that maintained proteomic while suppressing unwanted features in the image [7]. Once the noise is removed to enhance the contrast of the image, the imtophat filter is used to create with required edges. Since, the information of edges is needed for the proper segmentation. Then enhanced image is segmented using DBSCAN clustering with morphological operators discussed in the following section.

## 3. Extraction of Region of Interest (RoI)

The RoI is extracted using the DBSCAN clustering with morphological operators from thresholded image which is obtained by applying Local Adaptive Thresholding method in order to reduce the complexity of segmentation process. The schematic diagram of RoI extraction is shown in figure 2 and its detailed description is given subsequent sections.

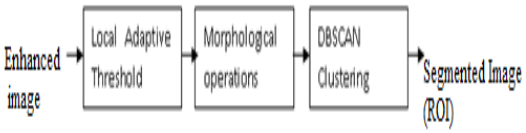



Figure 2: Schematic diagram for ROI

### 3.1 Local Adaptive Threshold

The enhanced image is thresholded by using local adaptive method. The mean value of the sub image is used as the threshold value for the pixel and works well in strict cases of the images that have approximately half to the pixels belonging to the objects and the other half to background. Threshold image g(x, y) can be defined:

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) < T \\ 0 & \text{otherwise} \end{cases}$$

The binary image obtained after thresholding contains several blocks holes. The morphological operators such as opening and closing are applied with large structuring element to isolate the object that is corresponding to the prostate region.

### 3.2 Morphological operators

Morphological operations are important in the digital image processing, since that can rigorously quantify many aspects of the geometrical structure of the way that agrees with the human intuition and perception [9, 10]. It emphasizes on studying geometry structure of image. The relationship between each part of image can be identified when processing image with morphological theory. Accordingly, the structural character of image in the morphological approach is analyzed in terms of some predetermined geometric shape known as structuring element [9]. The morphological operators opening and closing are defined as follows:

$$A \circ B = (A \ominus B) \oplus B$$
$$A \bullet B = (A \oplus B) \ominus B$$

where, A is an image and B is a structuring element. The

area must be isolated from other white regions in the thresholded image is a difficult task in TRUS prostate medical images.

### 3.3 DBSCAN Clustering

With the aim of separating background from TRUS image to target possible prostate, pixels of thresholded images are grouped by using DBSCAN. It takes a binary image, and delineates only significantly important regions. The expected outcome is desired boundary of the TRUS prostate image. The DBSCAN Algorithm for segmentation is detailed in figure 3. Technically, this algorithm is appropriate to tailor density based algorithm in which cluster definition guarantees that the number of positive pixels is equal to or greater than minimum number of pixels (MinPxl) in certain neighborhood of core points. The core point is that the neighborhood of a given radius (Eps) may contain at least a minimum number of positive pixels (MinPxl), i.e., the density in the neighborhood should exceed pre-defined threshold (MinPxl) [11, 12].

| DBSCAN Algorithm<br>INPUT: Enhanced TRUS prostate image<br>OUTPUT: Segmented image which contains only prostate |
|---|
| Step1: Set epsilon (Eps) and minimum points (MinPts).<br><br>Step2: Starts with an arbitrary starting point that has not been visited and then finds all the neighbor points within distance Eps of the starting point.<br><br>Step3: If the number of neighbors is greater than or equal to MinPts, a cluster is formed. The starting point and its neighbors are added to this cluster and the starting point is marked as visited. The algorithm then repeats the evaluation process for all the neighbors recursively.<br><br>Step4: If the number of neighbors is less than MinPts, the point is marked as noise. |

Figure 3: DBSCAN Algorithm

## 4. Feature Extraction

Texture is one of the most important characteristics of an image. It is used to describe the local spatial variations in image brightness which is related to image properties such as coarseness, and regularity. This is achieved by performing numerical manipulation of digitized images to get quantitative measurements. Texture analysis can potentially expand the visual skills of the expert eye by extracting image features that are relevant to diagnostic problem and not necessarily visually extractable. In order to capture the spatial dependence of gray-level values which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is discussed for texture consideration. Since texture shows its characteristics by both each pixel and pixel values. Normally texture analysis can be grouped into four categories: model-based, statistical-based, structural-based, and transform-based methods. Model-based methods are based on the concept of predicting pixel values based on a mathematical model. Statistical methods describe the image using pure numerical analysis of pixel intensity values. Structural approaches seek to understand the hierarchal structure of the image. Transform approaches generally perform some kind of modification to the image, obtaining a new "response" image that is then analyzed as a representative proxy for the original image [16].

This paper only focuses on statistical approaches, which represent texture with features that depend on relationships between the grey levels of the image. It is very helpful to know that different tissues have different textures [13]. Benign tumors are described as regular masses with homogenous internal echoes, while carcinomas are masses with fuzzy borders and heterogenous internal echoes. Statistical features are used in this paper where different texture features are constructed from the identified regions of interest of the TRUS images. In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. Depending on the number of pixels in each combination, statistics are classified into first-order, second-order and higher-order statistics. A typical TRUS image of prostate contains a vast amount of heterogeneous information that depicts different parts. In order to build a robust diagnostic system towards correctly classifying normal and abnormal, we have to present all the available information that exists in TRUS image to the diagnostic system so that it can easily discriminate between the normal and the abnormal tissue [14, 15]. In this paper, segmented image (ROI) is utilized to construct the feature sets using Histogram method, Gray-Level Run-Length Method (GLRLM), and Grey-Level Co-occurrence Matrix (GLCM). And then each features set and the various combinations of them used for the classification. In this paper we analyze TRUS images using three different texture extraction methods. Performance of the combinations of the above three methods are also analyzed.

### 4.1 Intensity Histogram Features

Intensity Histogram analysis has been extensively researched in the initial stages of development of this algorithm. The characteristic of the histogram has close relationship with the characteristic of image such as brightness and contrast. The characteristic of a histogram

and hence, the characteristic of an image can be expressed using the following measurements [17].

Mean reveals the general brightness of an image. Bright image should have high mean while dark image should have low mean, and also mean values characterize individual calcifications. Standard deviation or variance reveals the contrast of an image. Image with good contrast should have high variance. Standard Deviations (SD) also characterize the cluster. Skew measures is how asymmetry (unbalance) the distribution of the gray level. Image with bimodal histogram distribution (object in contrast background) should have high variance but low skew distribution (one peak at each side of mean). Energy measurement is closely related to skew. Highly skew distribution usually gives high-energy measurement. Entropy measures the average number of bits to code each gray level. It has inverse relationship with skew and energy measurement. Highly skew distribution tends to yield low Entropy. These are summarized in Table 1. Within ROI (ie segmented prostate region) a histogram distribution of the image is computed. Then six features are calculated for classification.

Table 1: Histogram Features

| S.No | Features | Formulae |
|---|---|---|
| 1 | Mean | $\sum_{i=1}^{N} ih(i)$ |
| 2 | Variance | $\sum_{i=1}^{N} (i-\mu)^2 h(i)$ |
| 3 | Skewness | $\sum_{i=1}^{N} (i-\mu)^3 h(i)$ |
| 4 | Kurtosis | $\sum_{i=1}^{N} [i-\mu)^4 h(i)] - 3$ |
| 5 | Entropy | $-\sum_{g=0}^{L-1} P(g) \log_2 [P(g)]$ |
| 6 | Energy | $\sum_{g=0}^{L-1} [P(g)]^2$ |

## 4.2. Gray Level Cooccurrence Matrix

The Gray Level Coocurrence Matrix (GLCM) method is a way of extracting second order statistical texture features [15, 18]. It models the relationships between pixels within the region by constructing Gray Level Co-occurrence Matrix. The GLCM is based on an estimation of the second-order joint conditional probability density functions p(i, j | d, θ) for various directions θ = 0, 45, 90, 135°, etc., and different distances, d = 1, 2, 3, 4, and 5. The function p(i, j | d, θ) is the probability that two pixels, which are located with an intersample distance d and a direction θ, have a gray level i and j. The spatial relationship is defined in terms of distance d and angle θ. If the texture is coarse, and distance d is small, the pair of pixels at distance d should have similar gray values. Conversely, for a fine texture, the pairs of pixels at distance d should often be quite different, so that the values in the GLCM should be spread out relatively uniformly [19, 20]. Similarly, if the texture is coarser in one direction than another, then the degree of spread of the values about the main diagonal in the GLCM should vary with the direction θ [21]. The figure 3 represents the formation of the GLCM of the grey-level (4 levels) image at the distance d = 1 and the direction θ = 0°.

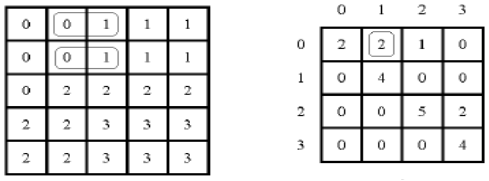



Figure 3(a): image with 4 grey level 3(b): GLCM for d = 1 and θ = 0°

The thin box in figure 3(a) represents pixel-intensity 0 with pixel intensity 1 as its neighbour in the direction θ = 0º. There are two occurrences of such pair of pixels. Therefore, the GLCM matrix formed with value 2 in row 0 and column 1. This process is repeated for other pair of intensity values. As a result, the pixel matrix represented in Figure 3(a) can be transformed into GLCM as shown in Figure 3(b). In addition to the direction (0º), GLCM can also be formed for the other directions 45º, 90º and 135º as shown in Figure 4.

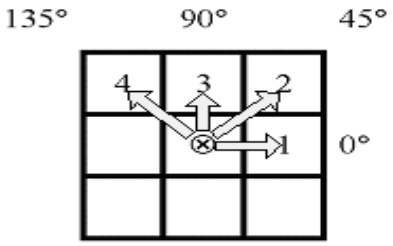



Figure 4: Direction of GLCM generation.

The pixels 1, 2, 3 and 4 are representing the directions (θ) 0º, 45º, 90º and 135º respectively for distance d = 1 from the pixel x.

## 4.3. Gray-Level Run-Length Matrix

Texture is understood as a pattern of grey intensity pixel in a particular direction from the reference pixels. Grey-Level Run-Length Matrix (GRLM) is a matrix from which

the texture features can be extracted for texture analysis [22]. It is a way of searching the image, always across a given direction, for runs of pixels having the same gray level value. Run length is the number of adjacent pixels that have the same grey intensity in a particular direction. Gray-level run-length matrix is a two-dimensional matrix where each element is the number of elements j with the intensity i, in the direction θ.

Thus, given a direction, the run-length matrix measures for each allowed gray level value how many times there are runs of, for example, 2 consecutive pixels with the same value. Next it does the same for 3 consecutive pixels, then for 4, 5 and so on. Note that many different run-length matrices may be computed for a single image, one for each chosen direction. The GLRLM is based on computing the number of gray level runs of various lengths [23]. A gray level run is a set of consecutive and collinear pixel points having the same gray level value. The length of the run is the number of pixel points in the run. The gray level run-length matrix is as follows.

$$R(\theta) = (g(i, j) \mid \theta),\ 0 \leqslant i \leqslant Ng,\ 0 \leqslant j \leqslant Rmax;$$

where Ng is the maximum gray level and Rmax is the maximum length. Figure 5 shows the sub image with 4 gray levels for constructing the GLRLM. Figure 6 shows that the GLRLM in the direction of 0 of the sub image in Figure 5.

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 1 | 3 | 4 | 4 |
| 3 | 2 | 2 | 2 |
| 4 | 1 | 4 | 1 |

Figure 5: Matrix of Image

| Gray Level | Run Length(j) | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 4 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 | 0 |
| 3 | 3 | 0 | 0 | 0 |
| 4 | 3 | 1 | 0 | 0 |

Figure 6: GLRL Matrix

In addition to the 0º direction, GLRLM can also be formed in the other direction, i.e. 45º, 90º or 135º.

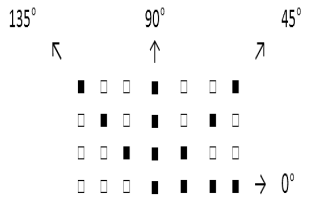



Figure 7: Run Direction

Table 1: GLRLM Features

| *S.No* | *Features* | *Formulae* |
|---|---|---|
| 1 | Short Run Emphasis (SRE) | $\frac{1}{n}\sum_{i,j}\frac{p(i,j)}{j^2}$ |
| 2 | Long Run Emphasis (LRE) | $\frac{1}{n}\sum_{i,j} j^2 p(i,j)$ |
| 3 | Grey Level Non-uniformity (GLN) | $\frac{1}{n}\sum_{i}\left(\sum_{j} p(i,j)\right)^2$ |
| 4 | Run Length Non-uniformity | $\frac{1}{n}\sum_{i}\left(\sum_{i} p(i,j)\right)^2$ |
| 5 | Run percentage (RP) | $\sum_{i,j}\frac{n}{p(i,j)\,j}$ |
| 6 | Low Grey Level Run Emphasis (LGRE) | $\frac{1}{n}\sum_{i,j}\frac{p(i,j)}{i^2}$ |
| 7 | High Grey Level Run Emphasis (HGRE) | $\frac{1}{n}\sum_{i,j} i^2 p(i,j)$ |

Seven texture features can be extracted from the GLRLM. These features uses grey level of pixel in sequence and is intended to distinguish the texture that has the same value of SRE and LRE but have differences in the distribution of gray levels. Once features sets are constructed using Histogram features, GLCM, GRLM, and their combination. Then the next section explicates SVM classifier for the classification of extracted features.

## 5. Classification

The classification of the image is the final step. It is a significant area of research and of practical applications in a variety of fields including pattern recognition, artificial intelligence medicine and vision analysis. The term image classification refers to the labeling of images into one of a number of predefined categories. For classification of images into class labels C i for i= 1 to m where m is the number of classes we use the feature set F, where F= $\{f_1, f_2, \ldots, f_n \mid f_i \cap f_j = \phi, i \neq j\}$ stands for the selected feature

set of cardinality n (n=|F|). For classification, the SVM is adapted and its detailed description is given below.

### 5.1 SVM Classifier

Support vector machine is based on statistical learning technique which is well-founded in modern statistical learning theory [27]. The Support Vector Machines were introduced by Vladimir Vapnik and his colleagues. The earliest mention was in (Vapnik, 1979). SVM is a useful technique for data classification [28]. It is also be a leading method for solving non-linear problem. A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. A step in SVM classification involves identification as which are intimately connected to the known classes [24].

Support vector machines use the training data to crate the optimal separating hyperplane between the classes. The optimal hyperplane maximizes the margin of the closest data points. A good separation is achieved by the hyperplane that has the largest distance to the nearest training features of any class (so-called functional margin). Maximum-margin hyperplane and margins for a SVM trained with samples from two classes. Samples on the margin are called the support vectors (Fatima Eddaoudi, and Fakhita Regragui,: 2011). SVM divides the given data into decision surface. Decision surface divides the data into two classes like a hyper plane. Training points are the supporting vector which defines the hyper plane. The basic theme of SVM is to maximize the margins between two classes of the hyper plane (Steve, 1998).

Basically, SVMs can only solve binary classification problems. They have then been extended to handle multi-class problems. The idea is to decompose the problem into many binary-class problems and then combine them to obtain the prediction. To allow for multi-class classification, SVM uses the one-against-one technique by fitting all binary sub classifiers and finding the correct class by a voting mechanism.

The "one-against-one" approach (Knerr et al., 1990) is implemented in SVM for multiclass classification. If K is the number of classes, then K(K - 1)/2 binary classifiers are constructed and trained to separate each pair of classes against each other, and uses a majority voting scheme (max-win strategy) to determine the output prediction. For training data from the i th and the j th classes, we solve the following two-class classification problem ( Metehan Makinacı: 2005). Let us denote the function associated with the SVM model of $\{c_i, c_j\}$ as:

$$g(x)_{ij} = sign(f(x)_{ij})$$

An unseen example, x, is then classified as:

$$f(x) = \arg\max_i \sum_{i=1}^{K} \sum_{j=1 \wedge i \neq j}^{K} V_{ij}(x)$$

where:

$$V_{i,j}(x) = \begin{cases} 1 \, if \;\; g_{ij}(x) = 1 \\ 0 \, if \; g_{ij}(x) = -1 \end{cases}$$

Each feature set is examined using the Support Vector Machine classifier. In classification, we use a voting strategy: each binary classification is considered to be a voting where votes can be cast for all data points x in the end point is designated to be in a class with the maximum number of votes. In case that two classes have identical votes, though it may not be a good strategy, now we simply choose the class appearing first in the array of storing class names.

The objective of any machine capable of learning is to achieve good generalization performance, given a finite amount of training data, by striking a balance between the goodness of fit attained on a given training dataset and the ability of the machine to achieve error-free recognition on other datasets. With this concept as the basis, support vector machines have proved to achieve good generalization performance with no prior knowledge of the data. The optimal separating hyperplane can be determined without any computations in the higher dimensional feature space by using kernel functions in the input space. Commonly used kernels include:

Linear Kernel: $K(x, y) = x.y$

Radial Basis Function (Gaussian) Kernel: $K(x, y) = e^{(-||x - y||^2 / 2\sigma^2)}$

Polynomial Kernel: $K(x, y) = (x.y + 1)^d$

The next section presents the experimental evaluation of the proposed methods for the task of classification.

## 6. EXPREMENT ANALYSIS AND DISCUSSION

Algorithms discussed in the previous sections have been implemented using MATLAB. The intact of 5500 images in normal and abnormal categories which are transformed into histogram, Gray Level Cooccurrence matrix (GLCM) and Grey Level Run-Length matrix (GLRLM). The nine histogram features, twenty two GLCM features and eleven GRLM features are extracted from the corresponding models respectively. The performance of various texture models are analyzed by using statistical parameters such as sensitivity, specificity, accuracy and discussed. The statistical parameters with formula are given in Table 3.

Table 3.Formula for Measures

| ***Measures*** | ***Formula*** |
|---|---|
| Sensitivity | TP/(TP+FN) |
| Specificity | TN/(TN+FP) |
| Accuracy | (TP+TN)/(TP+FP+TN+FN) |

TP- predicts cancer as cancer.
FP - predicts cancer as normal.
TN - predicts normal as normal.
FN - predicts normal as cancer

Sensitivity and Specificity are the two most important characteristics of a medical test. Sensitivity is the probability that the test procedure declares an affected individual affected (probability of a true positive). Specificity is the probability that the test procedure declares an unaffected individual unaffected (probability of a true negative). Accuracy measures the quality of the classification. It takes into account true and false positives and negatives. Accuracy is generally regarded with balanced measure whereas sensitivity deals with only positive cases and specificity deals with only negative cases. TP is number of true positives, FP is number of false positives, TN is number of true negatives and FN is number of false negatives. A confusion matrix provides information about actual and predicted cases produced by classification system. The performance of the system is examined by demonstrating correct and incorrect patterns. They are defined as confusion matrix in Table 4.

Table 4.Confusion Matrix

| Actual | Predicted | |
|---|---|---|
| | Positive | Negative |
| Positive | TP | FP |
| Negative | FN | TN |

The higher value of both sensitivity and specificity shows better performance of the system. The constructed feature sets are separately tested using the SVM classifier. The computational results are presented in Table5.

Table 5: computational results

| ***Methods*** | ***sensitivity*** | ***Specificity*** | ***Accuracy*** |
|---|---|---|---|
| Histogram | 0.81090909 | 0.9979591 | 0.82978333 |
| GLRLM | 0.82495575 | 1.0000000 | 0.85000000 |
| GLCM | 0.83790900 | 1.0000000 | 0.87895500 |
| GLRLM+HIST | 0.84615384 | 0.9992960 | 0.89683333 |
| GLCM+HIST | 0.84745762 | 1.0000000 | 0.90700000 |
| GLCM+GLRLM | 0.86206896 | 1.0000000 | 0.91333333 |
| ALL | 0.91743119 | 1.0000000 | 0.92833333 |

The results in Table 5 shows that the texture features extracted from different models discriminate malignant and benign with different accuracy. The individual performance measures are exposed in figure 5(a) – 5(c).

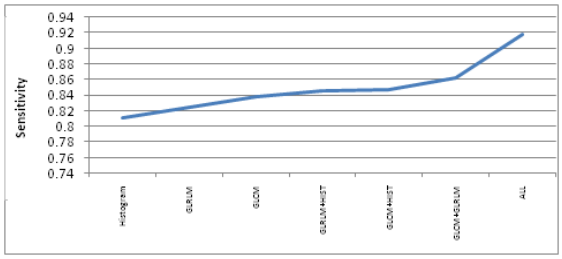



Figure 5(a): performance of sensitivity

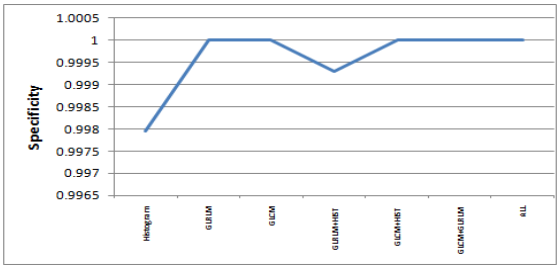



Figure 5(b): performance of specificity

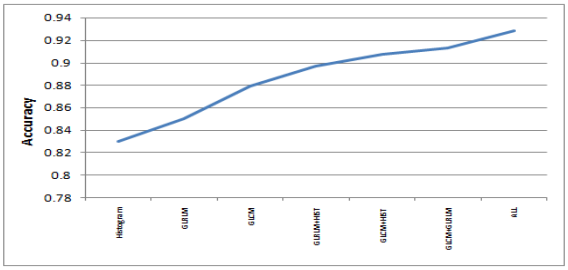



Figure 5(c): performance of accuracy

From the table 5, we observed that the maximum and minimum classification accuracies are 93% and 83% with SVM classifier. The histogram features discriminate between malignant masses and benign masses on TRUS prostate images with 83% accuracy, 81% sensitivity and 99% specificity levels that are relatively poorer compare to others. GLRLM features yielded an accuracy of 85% for distinguishing malignant and benign masses on TRUS prostate images. It is 2% higher than features based on histogram. The GLCM features achieved an accuracy of 88% where 84% sensitivity and 100% specificity. The accuracy of GLCM features 3% higher than GLRLM feature and 5% higher than Histogram features.

The combination of Histogram features and GLRLM features is achieved 90% of the accuracy, where as 91% of accuracy is produced when combined the Histogram features with GLRLM features. The accuracy difference between these two methods is only 1% even while the sensitivity and specificity of these two are almost same 85% and 100% respectively.The accuracy of 91 % is arrived by the combination of GLCM features and GLRLM features, whilst 86% sensitivity and 100% specificity. The predicted accuracy is 5%, which is 3% higher than GLRLM, GLCM respectively. The combination of Histogram features, GLCM features and GLRLM features produces the highest accuracy of 93%, 91% sensitivity, and 100% specificity followed by GLCM features combined with GLRLM features with the accuracy of 91%, 86% sensitivity, and 100% specificity.

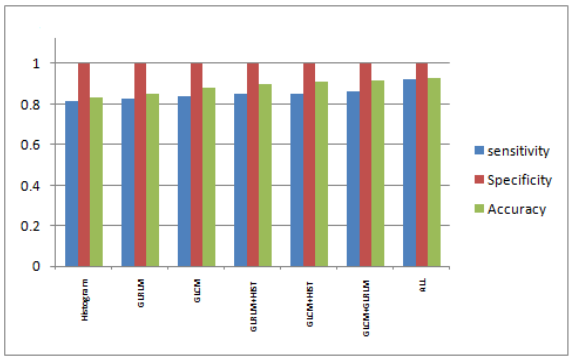



Figure 6: Relative performance measures

The Figure 6 shows that the relative performance measures with respect to proposed methods. By considering the different texture methods independently, it is not able to confirm that there is a universal method for best classification. However, usually the statistical cooccurrence (GLCM) features are used. The combination of various methods features produces a significant increase in the accuracy levels. It is interesting to note that using combined features produces relatively good classification results. The percentage of accuracy of combined features is higher than the values obtained from others. These analyses conclude that the combination of Histogram, GLCM and GLRLM texture features achieves best classification accuracy for distinguishing between malignant masses and benign masses on TRUS prostate images.

## 7. Conclusion

Texture analysis is a potentially valuable and versatile in TRUS imaging for prostate cancer interpretation. In some cases, radiologists face difficulties in directing the tumors. In this work, feature extraction methods for the TRUS prostate cancer classification problem and an innovative approach (combined features) of finding the malignant and benign masses from the TRUS prostate medical images are proposed and analyzed. The performances of classifiers for the texture-analysis methods are evaluated using various statistical parameters such as sensitivity, specificity and accuracy. The experiment results show that there is considerable performance variability among the various texture methods. The histogram features and GLRLM features performances are considerably poor. The combination histogram features, GLRLM features and GLCM features outperformed well in discriminating between or among prostate cancer. Using proper feature selection method accuracy may be improved efficiently in future.

## References

[1] Vidhi Rawat ,Alok jain, and Vibhakar shrimali: "Investigation and Assessment of Disorder of Ultrasound B-mode Images" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 2, February 2010, PP.289 – 293.

[2] Ransohoff DF, McNaughton Collins M, Fowler FJ. Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback. Am J Med. 2002 Dec 1;113(8):663-7. Review

[3] Grossfeld GD, Carroll PR. Prostate cancer early detection: a clinical perspective. Epidemiol Rev. 2001;23(1):173-80. Review.

[4] Cookson MM. Prostate cancer: screening and early detection. Cancer Control. 2001 Mar-Apr;8(2):133-40. Review

[5] Ferdinand Frauscher: "Contrast-enhanced Ultrasound in Prostate Cancer" Imaging and Radiotherapy, © T O U C H B R I E F I N G S 2 0 0 7. Pp 107-108.

[6] Ethan J Halpern, MD: "Contrast-Enhanced Ultrasound Imaging of Prostate Cancer",

[7] Thangavel, K., Manavalan, R. and Laurence Aroquiaraj . I. (2009), Removal of Speckle Noise from Ultrasound Medical Image based on Special Filters: Comparative Study, ICGST-GVIP Journal, ISSN 1687-398X, Volume (9), Issue (III), pp. 25-32.

[8] Sinha, G.R., Kavita Thakur and Kowar, M.K. (2008). "Speckle reduction in Ultrasound Image processing", Journal of Acoustical. Society of India, Vol. 35, No. 1, pp. 36-39.

[9] Anil K. Jain. (1989) Fundamentals of Digital Image Processing: Prentice Hall.

[10] Gonzalez, R. and Woods, R. (2002). Digital Image Processing. 3rd Edn., Prentice Hall Publications, pp. 50-51.

[11] Ester, M., Kriegel, H.P., Sander, J., and Xu, X.(1996) 'A density-based algorithm for discovering clusters in large spatial databases with noise'. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, Portland: AAAI Press. pp. 226-231.

[12] Sinan Kockara, Mutlu Mete, Bernard Chen, and Kemal Aydin:(2010). 'Analysis of density based and fuzzy c-means clustering methods on lesion border extraction in dermoscopy images', From Seventh Annual MCBIOS Conference: Bioinformatics Systems, Biology, Informatics and Computation Jonesboro, AR, USA. pp. 19 - 20.

[13] Amadasun, M. and King, R., (1989) 'Textural features corresponding to textural properties', IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 5, pp. 1264 - 1274.

[14] Haralick, R.M. , Shanmugan, K.. and Dinstein, I.( 1973) 'Textural Features for Image Classification', IEEE Tr. on Systems, Man, and Cybernetics, Vol SMC-3, No. 6, pp. 610-621.

[15] Haralick, R.M. (1979) 'Statistical and Structural Approaches to Texture', Proceedings of the IEEE, Vol. 67, No. 5, pp. 786-804.

[16] Mohamed, S.S. and Salama M.M. ( 2005 ) 'Computer Aided diagnosis for Prostate cancer using Support Vector Machine' Publication: Proc., medical imaging conference, California, SPIE Vol. 5744, pp. 899-907.

[17] Chitrakala, S. Shamini, P. and Manjula, D: "Multi-class Enhanced Image Mining of Heterogeneous Textual Images Using Multiple Image Features" , Advance Computing Conference, 2009. IACC 2009. IEEE International , 2009 , Page(s): 496 – 501.

[18] Haralick R. M., Shanmugam K., Dinstein I. Textural Features of Image Classification. IEEE Transactions on Systems, Man and Cybernetics. 1973. Vol. 3(6). P. 610-621.

[19] Soh L., Tsatsoulis C. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. IEEE Transactions on Geoscience and Remote Sensing. 1999. Vol. 37(2). P. 780 – 795.

[20] Clausi D A. An analysis of co-occurrence texture statistics as a function of grey level quantization. Canadian Journal of Remote Sensing. 2002. Vol. 28(1). P. 45-62.

[21] A. Sakalauskas, A. Lukosevicius1, and K. Lauckaite: "Texture analysis of transcranial sonographic images for Parkinson disease Diagnostics" ISSN 1392-2114 ULTRAGARSAS (ULTRASOUND), Vol. 66, No. 3, 2011. Pp 32 – 36.

[22] K.Thangavel, M.Karnan, R.Sivakumar and A. Kaja Mohideen " Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms" AIML Journal, Volume (5), Issue (3), pp.29-40., September, 2005.

[23] Jong Kook Kim, Jeong Mi Park, Koun Sik Song and Hyun Wook Park "Texture Analysis and Artificial Neural Network for Detection of Clustered Microcalcifications on Mammograms" IEEE, pp.199 – 206, 1997.

[24] Devendran V et. al., "Texture based Scene Categorization using Artificial Neural Networks and Support Vector Machines: A Comparative Study," ICGST-GVIP, Vol. 8, Issue IV, pp. 45-52, December 2008.

[25] L. Gagnon and F.D. Smaili: 'Speckle Noise Reduction of Airborne SAR Images with Symmetric Daubechies Wavelets', SPIE Proc. #2759, pp. 1424,1996.

[26] H. Guo, J E Odegard, M.Lang, R.A.Gopinath, I.W.Selesnick, and C.S. Burrus, "Wavelet based Speckle reduction with application to SAR based ATD/R", First Int'I Conf. on image processing , vol. 1, pp. 75-79,Nov 1994.

[27] Vapnik V. (1995). The Nature of Statistical Learning Theory, chapter 5. Springer-Verlag, New York.

[28] Vapnik.V, Statistical Learning Theory. John Wiley & Sons, New York, 1998.

**Manavalan Radhakrishnan** obtained M.Sc., Computer Science from St. Joseph's College of Bharathidasan University,Trichy, Tamilnadu, India, in the year 1999, and M.Phil., in Computer Science from Manonmaniam Sundaranar University, Thirunelveli, Tamilnadu, India in the year 2002. He works as Asst.Prof & Head, Department of Computer Science and Applications, KSR College of Arts and Science, Thiruchengode, Nammakal, Tamilnadu, India. He pursues Ph.D in Medical Image Processing. His areas of interest are Medical image processing and analysis, soft computing, pattern recognition and Theory of Computation.

**Thangavel Kuttiannan** received the Master of Science from Department of Mathematics, Bharathidasan University in 1986, and Master of Computer Applications Degree from Madurai Kamaraj University, India in 2001. He obtained his Ph. D. Degree from the Department of Mathematics, Gandhigram Rural University in 1999. Currently he is working as Professor and Head, Department of Computer Science, Periyar University, Salem. His areas of interests includes medical image processing, artificial intelligence, neural network, fuzzy logic, data mining, pattern recognition and mobile computing

# An Overview of Applications, Standards and Challenges in Futuristic Wireless Body Area Networks

**Ragesh G K[1], Dr.Baskaran K[2]**

**[1] Department of Computer Science and Engineering, Government College of Technology, Coimbatore-641013, Tamil Nadu, India**

**[2] Department of Computer Science and Engineering, Government College of Technology, Coimbatore-641013, Tamil Nadu, India**

**Abstract**

Recent technical advancements in low-power integrated circuits, ultra low-power RF (radio frequency) technology, wireless communications and micro sensors allowed the realization of Wireless Body Area Networks (WBANs). It is one of the latest technologies in health care diagnosis and management. A body area network wirelessly connects independent nodes (e.g. medical devices, earphones, sensors, actuators) attached to the body surface, implanted into tissues/body, or dispersed in the clothing for applications in home/health care, sports, entertainment, defense, ambient intelligence, pervasive computing and many other areas. These sensors offer promising applications in areas such as real time health monitoring, interactive gaming and consumer electronics. WBAN does not compel the patient to stay in the hospital thereby giving much physical mobility. Thus it greatly increases the efficiency of a health care system. This paper presents an overview on the various aspects of WBAN including sensors used, applications, power efficiency, communication protocols, security requirements, existing projects in WBANs and challenges faced in wireless body area networks.

Keywords: *Wireless body area network, WBAN, body area network, real time health monitoring, bio sensor networks, wearable sensors.*

## 1. Introduction

The growing cost of healthcare and the aging population in developed countries have introduced great challenges for governments, healthcare providers and healthcare industry. There is great interest in using emerging wireless technologies to support remote patient monitoring in an unobtrusive, reliable and cost effective manner thereby providing personalized sustainable services to patients. Wireless Body Area Networks (WBANs) is one such emerging technology that has the potential to significantly improve health care delivery, diagnostic monitoring, disease-tracking and related medical procedures. A crucial aspect of WBANs is their ability to provide highly reliable communications for medical devices, especially those implanted in the human body. Wireless Body Area Network (WBAN) consists of a number of inexpensive, lightweight, miniature sensors which could be located on the body as tiny intelligent patches, integrated in to clothing or implanted beneath the skin or embedded deeply in to the body tissues. Their main purpose is to enable doctors and other medical staff to safely monitor the health status of patients. This WBAN technology brings affordable and efficient healthcare solutions to people that will improve their quality of life.

Strategically placed wearable or implanted wireless sensor nodes consistently monitor the patient's vital signs, such as electro cardiogram (ECG), EEG and blood pressure; or important environmental parameters like temperature and humidity. The patient related data (gathered data) from all WBANs may ultimately be sent to a centralized healthcare repository for permanent records. Physicians can remotely access this data to assess the state of health of the patient. Additionally the patient can be alerted using SMS, alarm, or reminder messages. In this article we present a survey of the state of the art in Wireless Body Area Networks. Our aim is to provide a better understanding of the current research issues in this emerging field. The remainder of this paper is organized as follows. First, the sensors of WBAN are discussed in Section 1. Next, the applications of BANs and energy efficiency of WBAN are discussed in Section 2 and 3. Section 4 gives an overview of the standardized technologies used for WBAN communication. Section 5 deals with the necessity of security in WBAN. Section 6 and 7 deals with Physical and MAC layers. Section 8 discusses the WBAN specific routing protocols and other protocols related to WBAN. Relation to wireless sensor networks and WBAN challenges is treated in section 9 and 10. An overview of existing projects and social issues of

WBAN is given in Section 11 and 12 respectively. Finally section 13 concludes the paper.

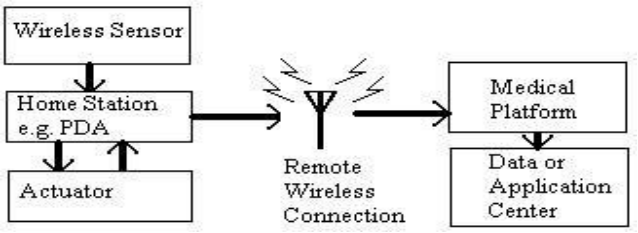



Fig.1 Data flow in a typical medical BAN

## 2. Sensors of WBAN

The sensors of a Body Area Network are extremely compact and complex in design. The fact that the sensors are so minute means that the patients will be able to lead a normal life, as the sensor devices are very unobtrusive. All sensors produced will contain the same basic elements such as a power supply and wireless transceiver as well as a control mechanism, a sensor and the casing that will hold all of the components together. The sensors will be designed in a way that allows them to be self-governing for the entire lifetime. BAN's work through a process of data being transmitted from an implanted device to an external device. The sensor which is implanted inside a patient's body interacts with other sensors and actuators wirelessly. The mechanism by which an agent acts upon an environment is known as an actuator. Artificial intelligent agent or any other autonomous being (human being or an animal) can be an agent. The Body Area Network functions by passing data from each sensor to a main station. The main station then fuses the data passed from each of the sensors and it is then sent to a recipient via the internet.

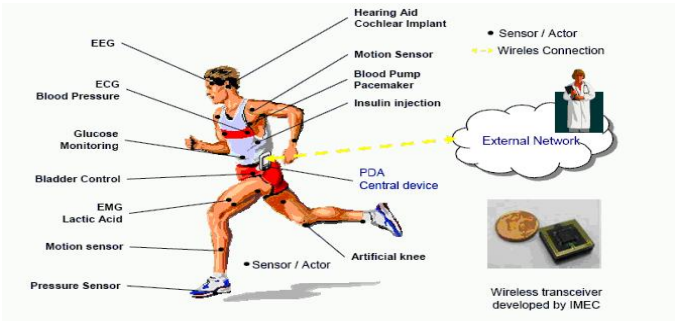



Fig.2 A BAN on an Athlete (Latré, 2005)

## 2. Applications of BANs

Due to the diverse components that can be connected and integrated, body area networks will be able to provide various functions in healthcare, emergency, work, research, lifestyle, sports, or military.

### 2.1 Medical Applications

BANs can provide interfaces for diagnostics, for remote monitoring of human physiological data, for administration of drugs in hospitals and as an aid to rehabilitation. In the future it will be possible to monitor patients continuously and give the necessary medication whether they are at home, in a hospital or elsewhere. Patients will no longer need to be connected to large machines in order to be monitored.

### 2.2 Lifestyle and Sports

BANs enable new services and functions for wireless body-centric networks including wearable entertainment system (e.g., music entertainment), navigation support in the car or while walking, museum or city guide, heart rate and performance monitoring in sports, infant monitoring, wireless cash card (e.g., display of recent transactions and checking of balance, etc).

### 2.3 Military Applications

The opportunities for using BANs in the military are numerous. Some of the military applications for BANs include monitoring health, location, temperature and hydration levels. A battle dress uniform integrated with a BAN may become a wearable electronic network that connects devices such as life support sensors, cameras, RF and personal PDAs, health monitoring GPS, and transports data to and from the soldier's wearable computer. The network could perform functions such as chemical detection, identification to prevent casualties from friendly fire and monitoring of a soldier's physiological condition. Calling for support, his radio sends and receives signals with an antenna blended into his uniform. As a result, BANs provide new opportunities for battlefield lethality and survivability.

## 3. Energy efficiency

Major sources of energy waste usually considered are: idle listening, over hearing and protocol overhead. A node is idle listening when it expects to receive packets and no packets are received. Listening is an expensive operation, and this should be avoided. A node is overhearing when it receives irrelevant packets. This wastes energy, because receiving is an expensive operation as well, nodes should only receive relevant packets. Protocol overhead is the ratio of the amount of bits needed to transmit a data packet over

the amount of bits of data packet itself. This will always be present, but should be kept to a minimum.

## 4. Technologies and Standards

A number of standardized technologies are related to WBAN research.

### 4.1 IEEE802.15.6

The IEEE 802.15 task group 6(BAN) is developing communication standard optimized for low power devices to provide a variety of applications including remote health care monitoring, consumer electronics, interactive gaming and other. Existing ISM bands as well as frequency bands approved by national medical and/or regulatory authorities can be used by this standard. The standard requires support for Quality of Service (QoS), extremely low power, and data rates up to 10 Mbps.

### 4.2 IEEE 802.15.4

Some researchers consider this as a MAC protocol and lots of research focuses on this protocol. However, research points out that the performance of IEEE802.15.4 is not sufficient for WBANs. The performance of this protocol in a multi hop environment is very poor.

### 4.3 Bluetooth

Bluetooth is a broadly available WPAN protocol and is very popular for current medical care solutions, especially because of the large range of available hardware implementations. However Bluetooth and other WPAN protocols have been designed for high data rate networks and large battery capacity which does not match the WBAN requirements. Also lowering the data rates will increase the protocol overhead.

## 5. Security Requirements

The use of wireless technology, especially to deliver health care, also brings with it a host of concerns about security and privacy. The security mechanism of the system is responsible for providing the following security services on specified biomedical data when requested to do so by the applications.

Data Encryption—the data is encrypted so that it is not disclosed whilst in transit. Data encryption service provides confidentiality against eavesdropping attacks.

Data Integrality—Data integrality service consists of data integrity and data origin authentication. Proper data integrity mechanisms at the BN and the BNC ensure that the received data is not altered by an adversary.

Freshness Protection—Data freshness ensures that the data frames are in order and are not reused.

Authentication— an efficient method against impersonation attacks.

## 6. Physical layer

A lot of research has investigated to physical layer. At the beginning of WBAN research a number of authors proposed Ultra Wide Band (UWB) as a physical layer for WBANs.UWB has the advantage of low energy consumption, good co-operation with existing wireless networks and a range large enough to support the entire body. Due to standardization issues and difficulties delivering the very high speeds.UWB does not progress well. As opposed to the wide bands proposed by UWB, other researchers propose the small, Industrial, Scientific and Medical (ISM) bands of the IEEE 802.15.4 and IEEE802.15.6.Current most working WBAN prototypes are based on ISM bands.

## 7. MAC layer

A number of WBAN specific MAC protocols exist. These can be divided in to single hop and multi hop protocols. The latter refers to the protocols which are optimized for multi hop topologies. The first protocols were designed based on a single hop topology. An example for this is Heart Beat driven MAC (H-MAC), which uses the heart beat to synchronize nodes. The protocol is specifically designed for WBANs; however traffic adaptations is not possible. Few other protocols have been developed usually IEEE 802.15.4 is preferred. Because of the dynamic nature, ad hoc network protocol could also be considered as WBAN protocols. Ad hoc network protocols are based on always –on radios, which matters their application to WBAN unfeasible.

## 8. WBAN specific routing protocols

When considering wireless transmission around and on the body, important issues are radiation absorption and heating effects on the human body. To avoid the heat generation, five thermal aware routing protocols were proposed. To reduce tissue heating, the radio's transmission power can be limited or traffic control algorithms can be used. Researchers showed that the bio effects caused by radio

frequency radiation are highly related to the incident power density, network traffic and tissue characteristics. A price-based rate allocation algorithm further shows that the bio effects can be reduced via power scheduling and traffic control algorithms. The Thermal Aware Routing Algorithm (TARA) routes data away from high temperature areas due to focusing data communications, defined as hotspots. When the temperature of a neighboring node is above a certain threshold, i.e., the node is becoming a hot spot, the packets will no longer be forwarded to the node but will be withdrawn and rerouted through alternate paths. The algorithm leads to a better temperature distribution over all the nodes in the network. However, TARA only considers the temperature as a metric. Consequently, it suffers from low network lifetime, a high ratio of dropped packets and low reliability, which is problematic for a WBAN. Improvements of TARA are Least Temperature Routing (LTR) and Adaptive Least Temperature Routing (ALTR). Unlike TARA, LTR always chooses the neighboring node with the lowest temperature as the next hop for routing. In order to maintain the network bandwidth, a predefined maximum hop count is used. When the number of hops exceeds this maximum, the packet is discarded. Loops are avoided by maintaining a list in the packet with the recently visited nodes. Unlike TARA, LTR always chooses the neighboring node with the lowest temperature as the next hop for routing. In order to maintain the network bandwidth, a predefined maximum hop count is used. When the number of hops exceeds this maximum, the packet is discarded. Loops are avoided by maintaining a list in the packet with the recently visited nodes. In general, temperature routing can be considered as a specific case of weight based routing. Results are promising, but reliability and energy efficiency can be hard to guarantee.

### 8.1 Other Routing Protocols

Similar to the related MAC protocols, a number of routing protocols for sensor and ad hoc networks could be considered good candidates for WBANs. The WSN protocols will focus on networks of a much larger scale while ad hoc network routing protocols will assume nodes with a larger battery and an always on radio.

## 9. Relation to Wireless Sensor Networks

In several papers, WBANs are considered to be a special type of WSN or Wireless Sensor and Actuator Network (WSAN) with its own requirements. However, traditional sensor networks do not tackle the specific challenges associated with human body monitoring. The most important difference is the need for reliable communication with each WBAN node, as opposed to the redundant character of WSN nodes. This corresponds to the typical medical application of WBANs, where only a single sensor per vital parameter is used. Moreover, the scale of WBANs is very small compared to typical large scale deployments of WSNs. In a WBAN, up to twenty nodes are expected to be deployed on a single person, while WSN protocols are usually designed for hundreds of nodes deployed in areas with diameters of hundreds of meters. A lot of research is being done toward energy efficient routing in ad hoc networks and WSNs. However, the proposed solutions are inadequate for WBANs. For example, in WSNs maximal throughput and minimal routing overhead are considered to be more important than minimal energy consumption. Energy efficient ad hoc network protocols only attempt to find routes in the network that minimize energy consumption in terminals with small energy resources, thereby neglecting parameters such as the amount of operations (measurements, data processing, access to memory) and energy required to transmit and receive a useful bit over the wireless link.

The following illustrates some main differences between Wireless Body Area Networks and Wireless Sensor Networks:

There are no redundant devices in WBANs in- spite of WSNs. All nodes in the network must be highly robust, reliable, and accurate. The lost information from one node often cannot be recovered by other nodes.

Because of the special features of the environment in which the WBAN operates (human body) the data loss is more significant. The signals of the sensors, specially the implanted ones, are considerably attenuated because the propagation of the waves takes place in or on a very lossy medium. Proprietary mechanisms may be required to ensure the QoS and real time data interrogation capabilities. However, in WSNs the data loss may be covered by other sensors.

The sensors which are either implanted into a tissue or attached on the surface of body must be very small in size to support unobtrusive monitoring of the patients. However, in WSNs the sensor size is not the main concern though smaller sensors are preferred. The small size of the WBAN sensors severely affects the power resources of the devices. The power supply recharge of the devices is often impossible. Thus, a long lifetime of the sensors is required.

The sensors in a WBAN are located in or on the human body which can be in motion. This challenge for WBAN is rarely available for WSNs. Thus the WBAN must be robust against the high probable network topology changes. In addition, biological variation and complexity cause a more variable structure.

## 10. WBAN Challenges

| Challenges | WBAN |
|---|---|
| Scale | As large as human body parts(millimeters/centimeters) |
| Node Number | Fewer, more accurate sensors nodes required (limited by space) |
| Node Function | Single sensors, each performs multiple tasks |
| Node Accuracy | Limited node number with each required to be robust and accurate |
| Node Size | Pervasive monitoring and need for miniaturization |
| Data Protection | High level wireless data transfer security required to protect patient's information |
| Access | Implantable sensor replacement difficult and requires biodegradability |
| Bio Compatibility | A must for implantable and some external sensors. Likely to increase cost |
| Context Awareness | Very important because body physiology is very sensitive to context change |
| Wireless Technology | Low power wireless required, with signal detection more challenging |
| Data Transfer | Loss of data more significant, and may require additional measures to ensure QoS and real-time data interrogation capabilities |

Table.1 Challenges faced by WBAN

## 11. Existing WBAN Projects

In the recent years a lot of work related to WBANs has appeared in the literature. The attempts are mostly focused on proposing solutions for the issues of the WBANs. Before introducing the IEEE 802.15.6 standard by the IEEE 802.15 Working Group the structure of WBANs and protocols and mechanisms of the physical layer and MAC sub layer of WBANs have been one of the most important concerns which attracted attention of many researchers. There are currently several research groups throughout the world which focus on design and implementation of a WBAN. The researchers have employed different wireless technologies in their projects in the field of wireless short-range connectivity, such as the IEEE 802 family of WPANs, WLANs, Bluetooth and Zigbee. Due to major drawbacks of other WPAN and WLAN solutions the IEEE 802.15.4/ Zigbee system has been the most favoured approach in the existing projects before the IEEE 802.15.6 standard is introduced.

In [4] proposes a system that could perform real-time monitoring of complex conditions on streaming data from various body sensors within a Wireless Body Area Network (WBAN). The system enables personal medical applications to be developed using personal electronic devices combined together with sensors in a WBAN. The main techniques developed are the query language which supports windowing capabilities, and the query index using an Interval Skip List data structure with windowing support.

In[6] *Stevan Marinkovic and Emanuel Popovici* developed implemented and tested a Nano power Wake Up Radio mainly intended for Wireless Body Area Networks (WBANs), but it can be also used in other types of low power wireless networks. The radio was tested for power consumption and robustness to communication interferences from a wireless device commonly found around the person carrying a WBAN.

*Janani.K, V.R.SarmaDhulipala and R.M.Chandrasekaran* developed a *WSN based frame work for human health monitoring in* [7].In this paper the framework they proposed provides a clear understanding how WSN is used for remote monitoring of the patient's health. The paper mainly focus on the understandability of the remote patient monitoring done in hospital, the vital network parameters to be considered, scalability and power consumption.

*Jae-Hoon Choi, Heung-Gyoon Ryu of* Chungbuk *National University, Korea proposed a* new QAPM (*Quadrature-Amplitude-Position-Modulation*) scheme for improving power efficiency in [8]. In this paper, they were analyzed existing PSSK and new propose QAPM scheme. The PSSK and QAPM scheme are extension method for increase power efficiency. And the simulation results, shows that BER performance of QAPM and PSSK better than QAM and PSK in AWGN channel. Also throughput of QAPM has better throughput characteristics in low SNR than PSK, QAM and PSSK.

In Opportunistic Routing for Body Area Network [9] provides an opportunistic scheme to exploit the body movements during the walking to increase the life time of the network. In this work they exploited the motion of the body parts to increase the lifetime of the network. To evaluate the performance of the proposed scheme, the energy consumption of the network per bit for the single hop, multi-hop using relay node and the opportunistic scheme are compared. The results shows the proposed scheme can increase the life time of the network by decreasing the energy consumption in both the sensor and relay nodes while maintaining the same BER as the other two schemes.

In *Wearable ECG Monitor* project [10] a wearable ubiquitous healthcare monitoring system using integrated electrocardiogram (ECG), Photoplethysmogratphy (PPG), Skin Temperature and Accelerometer etc. were designed and developed. In this design, nonintrusive healthcare system was designed based on WBAN for wide area coverage with minimum battery power to support RF transmission. In this system, WBAN, Zigbee, is used to communicate between wearable physiological signal devices and the personalized mobile system. We have developed various devices such as a wearable chest, wrist and necklace Device. The wearable ubiquitous healthcare monitoring system allows physiological data to be transmitted in wireless sensor network for Mobile network.

In [11] *Mrinmoy Barua, M.S. Alam, Xiaohui Liang, and Xuemin (Sherman) Shen* an efficient secure data transmission scheme in WBAN is proposed with data integrity. The scheme is user-centric and the secure key is shared among all sensors in a WBAN to minimize any additional memory and processing power requirements. Security analysis and numerical results demonstrates that the scheme can minimize the mean waiting time of a real-time traffic in WBAN and provide proper security and privacy.

*Xigang Huang, Hangguan Shan,* and *Xuemin (Sherman) Shen* investigate the energy efficiency of cooperative communications in wireless body area network (WBAN) in [12]. Three transmission schemes had been compared in this paper. Direct transmission, single-relay cooperation, and multi-relay cooperation. For each of them, they analyzed its outage performance and studied the problem of optimal power allocation with the constraint of targeted outage probability.

## 12. Social Issues

A number of issues exist regarding the creation of BANs include system design issues and human issues. System Design Issues include (Jovanov, 2005):

- Sensor Types

What type of sensor should be included in the BAN? This will depend on where it is to be used and for what purpose

- Power Sources

If the BAN is designed to be used for a long period of time then the power sources must be appropriate. If it is going to be used for some short intense activity then a different source could be used.

- Wireless Communication Range

Is the person using the BAN likely to remain within a particular area? e.g.: a hospital, or are they likely to be outdoors? E.g.: a soldier in the desert.

- Sensor location and mounting

Could they be woven into the uniform of a solider or might they need to be small unobtrusive implants in the skin?

- Weight and size of sensor

If the person is confined to bed at home then the sensors could be of a different type from those used on a runner.

## 13. Conclusion

In this paper various key aspects of WBAN including sensors used, application areas, technologies and standards, routing protocols, WBAN challenges and existing WBAN projects are outlined. Also discussed energy requirements, security requirements and issues present in various layers of WBAN. Finally some of the social issues related to WBAN application are mentioned. There are many challenges that still need to be addressed, especially on high bandwidth and energy efficient communication protocols, interoperability between BANs and other wireless technologies, and the design of successful applications. Future work will be concentrating on the design of a context aware mechanism which will carefully optimize security, safety and usability.

## References

[1]Maulin Patel and Jian Feng Wang, "Applications, Challenges and Prospective In Emerging Body Area Networking Technologies", IEEE Wireless Communications, IEEE, 2010,pp. 1536-1284.

[2] Ming Li and Wenjing Lou "Data security and Privacy In wireless body area networks", IEEE Wireless Communications, February 2010,pp. 1536-1284 .

[3] Xigang Huangy, Hangguan Shany, and Xuemin Shen, "On Energy Efficiency of Cooperative Communications in Wireless Body Area Networks", IEEE,2011.

[4] Minsoo Lee, Okju Choi, Seul-A Lee and Nory Chou, "Real time monitoring of complex window conditions on Body sensor data for wireless body area networks",IEEE International Conference on Consumer Electronics (ICCE), 2011.

[5]Saeed Rashwand, Jelena Misic and Hamzeh Khazaei, "Performance Analysis of IEEE 802.15.6 under saturation condition and error prone channel," IEEE. WCNC-Network, 2011.

[6]Stevan Marinkovic and Emanuel Popovici "Nano- Power Wake-Up Radio Circuit for Wireless Body Area Network", IEEE,2011.

[7] Janani.K, V.R.SarmaDhulipala and R.M .Chandra Sekaran "A WSN based frame work for human health monitoring", International Conference on devices and communication, IEEE, 2011.
[8] Jae-Hoon Choi and Heung-Gyoon Ryu "A QAPM Quadrature Amplitude-Position-Modulation) scheme for improving power efficiency", IEEE, 2011.
[ 9] Arash Maskooki, Cheong Boon Soh, Erry Gunawan and Kay Soon Low , "Opportunistic Routing for Body Area Network", 1st IEEE International Workshop on Consumer eHealth Platforms, Services and Applications, IEEE,2011.
[10] Youngsung Kim, Il-yeon Cho, "Wearable ECG Monitor-Evaluation and Experimental Analysis", IEEE,2011.
[11] Mrinmoy Barua, M.S. Alam, Xiaohui Liang, and Xuemin (Sherman) Shen, "Secure and Quality of Service Assurance Scheduling Scheme for WBAN with Application to eHealth", IEEE WCNC 2011 –Network,2011.
[12] Xigang Huang, Hangguan Shan, and Xuemin "On Energy Efficiency of Cooperative Communications in Wireless Body Area Networks‖, IEEE WCNC 2011 –Network,2011.
[13] D. Cypher, N. Chevrollier, N. Montavont, and N. Golmie-"Prevailing over wires in healthcare environments: benefits and challenges," IEEE Communications Magazine, vol. 44, no. 4, Apr. 2006. pp. 56{63.
[14] S. Drude-"Requirements and application scenarios for body area networks," in Mobile and Wireless Communications summit 2007. 16th IST, Budapest, Hungary, Jul. 2007, pp. 1{5.
[15] M.A. Ameen, Ahsanun Nessa and Kyung Sup Kwak, "QoS issues with focus on Wireless Body Area Networks", Third 2008 International Conference on Convergence and Hybrid Information Technology, IEEE,2008.
[16] M.Umar Talha and Jahanzeb Ahmad, "Body Area Networks (BANS) – An Overview With Smart Sensors Based Telemedical Monitoring System",Iraq university, Karachi.
[17] Latré, B. Moerman, I., Dhoedt, B, Demeester, P. "Networking in Wireless Body Area Networks", http://www.ibcn.intec.ugent.be/css_design/research/topics/2005/Networking/Wireless/Body/Are/Networks.pdf, 2005.

**Ragesh G K**
Ragesh G K obtained his B.E degree in Electronics and Communication Engineering from Anna University, Chennai, Tamil Nadu, India in 2005 and his M.E degree in Communication systems in 2008 from Anna University, Chennai, Tamil Nadu, India. Currently he is a Research scholar in Department of computer science and engineering at Government College of Technology, Coimbatore, Tamil Nadu, India. His research interests include Bio sensor networks, wireless sensor networks, Mobile ad hoc networks and Network security.

**Dr.K.Baskaran**
Dr.K.Baskaran is a member of the IEEE and ISTE, obtained his PhD degree from Anna University, Chennai. Tamil Nadu, India. He is having 18 years of teaching experience and got several awards including Best System Paper Award from IETE in 2010.Currently he is working as an Associate Professor in Department of computer science and engineering at Government College of Technology, Coimbatore, Tamil Nadu, India. His research interests include MANETs, Wireless sensor networks, Cloud Computing, Wireless Body Area Networks and Network security.

# Secure Routing in Wireless Sensor Networks

**Mrs Soumyashree Sahoo [1], Mr Pradipta Kumar Mishra [2] and Prof.Dr.Rabi Narayan Satpathy[3]**

[1]**Department of Computer Science and Engineering, Biju Patnaik University of Technology, Hi-Tech Institute of Technology**
**Bhubaneswar, Orissa 752055/7, India**

[2] **Department of Computer Science and Engineering , Biju Patnaik University of Technology, Hi-Tech Institute of Technology**
**Bhubaneswar, Orissa 752055/7, India**

[3] **Department of Computer Science and Engineering , Biju Patnaik University of Technology, Hi-Tech Institute of Technology**
**Bhubaneswar, Orissa 752055/7, India**

**Abstract**

**Wireless sensor networks is the new concept in the field of networks consists of small, large number of sensing nodes which is having the sensing, computational and transmission power. Due to lack of tamper-resistant infrastructure and the insecure nature of wireless communication channels, these networks are vulnerable to internal and external attacks. Key Management is a major challenge to achieve security in wireless sensor networks. Key management includes the process of key setup, the initial distribution of keys and keys revocation. To provide security and proper routing or communication should be encrypted and authenticated. It is not easy to achieve secure key establishment without public key cryptography. In this thesis, some key management schemes have been purposed which will be valuable for secure routing between different sensor nodes.**

***Keywords: Security, Design, Sensor networks, Key management, Algorithms***

## 1. Introduction

Wireless sensor networks consist of spatially distributed sensors to co-operatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants. When sensor networks are deployed in a hostile environment, security becomes extremely important, as they are prone to different types of malicious attacks. For example, an adversary can easily listen to the traffic, impersonate one of the network nodes, or intentionally provide misleading information to other nodes. The flexibility, fault tolerance, high sensing fidelity, low-cost and rapid deployment characteristics of sensor networks create many new and exciting application areas for remote sensing.

Key Management includes the processes of key setup, the initial distribution of keys and key revocation. To provide security and communication should be encrypted and authenticated. The open problem is how to bootstrap secure communications between sensor nodes, i.e. how to set up secret keys between communicating nodes. This problem is known as the key agreement problem. There are three types of general key agreement schemes: trusted-server scheme, self-enforcing scheme, and key pre-distribution scheme. The trusted-server scheme is not suitable for sensor networks because there is no trusted infrastructure in sensor networks. The self-enforcing scheme that depends on asymmetric cryptography is inapplicable for using public-key algorithms due to limited computation and energy resources of sensor nodes. The third type of key agreement scheme is key pre-distribution, where key information is distributed among all sensor nodes prior to deployment. If we know, which nodes will be in the same neighbourhood before deployment, keys can be decided a priori. However, most sensor network deployments are random; thus, such a priori knowledge does not exist. Thus, we have to make tradeoffs between security and the available resources while designing an

efficient key management scheme to achieve better security. The key management schemes must establish a key between all sensor nodes that must exchange data securely, node addition or deletion should be supported, Security services such as authentication and key management are critical to secure the communication between sensors in hostile environments. The key management schemes that we are going to present, must establish a key between all sensor nodes that must exchange data securely, node addition or deletion should be supported, It should work in undefined deployment environment, unauthorized nodes should not be allowed to establish communication with network nodes. The proposed key management protocols have been presented in detail.

1. When a sensor node wants to communicate with another sensor node, they need to be authenticated with each other first. Motivated by this, we proposed an Authentication Scheme without storing direct key information in the memory of individual sensor nodes.
2. Two communicating nodes need to have a pair wise key for secure communication. We proposed a pair wise key establishment scheme by using a counter.
3. Compromised nodes could deviate the network behaviour by injecting false data or modifying data of correct nodes. Thus, compromised nodes should not be taking part in the group communication. To solve this problem we proposed the key management scheme for node revocation.

## 2. Related Work

All key management schemes can be categorized in to two types i.e. pre-distribution key management schemes where key information is distributed among all sensor nodes prior to deployment and in Insitu key management schemes which does not require keying information prior to deployment .This key pre-distribution can be divided into several categories based on Key Pool, Random Pair Wise Key, Key Space, Group, grid-based etc. In key pool based scheme, a large key pool is computed offline and each sensor is preloaded with keys Selected randomly without replacement from the key pool. These keys form a sensor's key ring. Laurent Eschenauer and Virgil D. Gligor proposed a probabilistic key pre distribution scheme[2]. The main idea was to let each sensor node randomly pick a set of keys from a key pool before deployment so any two-sensor nodes sharing at least one common key. Donggang Liu and Peng Ning [9] presented a key management scheme where sensors can be added dynamically without having to contact the previously deployed sensors. This scheme allows the network to grow. In 2003 Chan et al proposed the Random pair wise scheme [3]. The random pair wise scheme possesses perfect resilience against node capture attacks as well as support for node based revocation and resistance to node replication. In 2009 Stephen Anokye et al [6] made improvements on Chan et al's Random pair wise scheme. This scheme significantly reduces the memory and computational overheads and it has better connectivity. In that year Du et al proposed a pair wise key predistribution scheme that is scalable and flexible. It is substantially more resilient against node capture. In 2005 Zhen Yu and Yong Guan proposed a group based key pre distribution scheme [11] using sensor deployment knowledge where a sensor field is partitioned into hexagonal grids. It consists of a series of methods depending on how to distribute secret information among neighbouring groups and how much information to be stored in each node which achieves a higher degree of connectivity of the sensor network with a lower memory requirement and offers a stronger resilience against node capture attacks. In 2006 R. Kalindi et al proposed a Grid based Key Pre-Distribution Scheme [12] for Distributed Sensor Networks that uses multiple mappings of keys to nodes. In each mapping, every node gets distinct set of keys, which it shares, with different nodes. The key assignment is done such that, there will be keys in common between nodes in different sub-grids. After randomly being deployed, the nodes discover common keys, authenticate and communicate securely. This scheme is able to achieve better security. Being motivated by the above key management schemes, some protocols have been developed for solving some of the important security issues that are encountered in Wireless sensor networks.

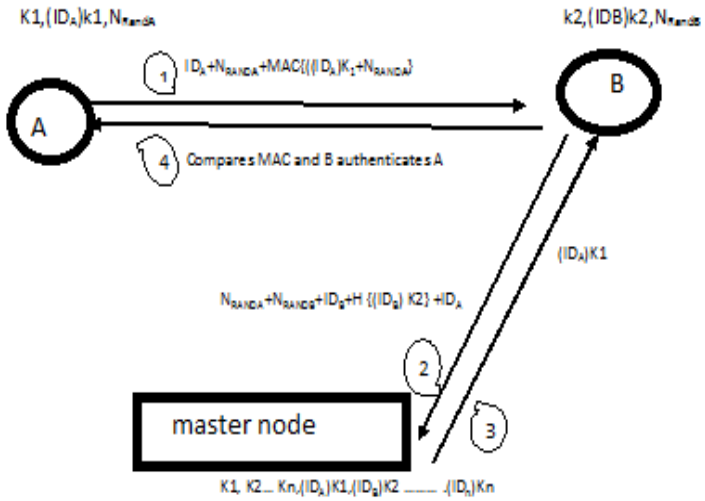
## 3. Proposed Key Management Schemes

Conventional key exchange and key distribution protocols based on infrastructures using trusted third parties are impractical for large-scale sensor networks because of the unknown network topology prior to deployment, communication range limitations, intermittent sensor-node operation, and network dynamics. To date, the only practical options for the distribution of keys to sensor nodes of large-scale distributed sensor networks whose physical topology is unknown prior to deployment would have to rely on key pre distribution. Keys would have to be installed in sensor nodes to accommodate secure connectivity between nodes. Security services such as authentication and key management are critical to secure the communication between sensors in hostile environments. Here three schemes have been proposed i.e. authentication, pair wise key establishment and a scheme for node revocation in an existing sensor network. The proposed key management schemes are presented in detail.

### 3.1 Authentication Scheme between two Individual Sensor Nodes:

When a sensor node wants to communicate with another sensor node, they need to be authenticated with each other first. In this scheme, it is assumed that no key information will be stored in individual sensor nodes; instead, every key information will be stored at a high-end sensor node i.e. the master node. By protecting only that master node, we can make the entire sensor network secure. The scheme is shown through an algorithm explained below.

1. Assuming the case of sensor node A and sensor node B in the network wants to securely communicate with each other.
2. The master node must contain keys of all sensor nodes i.e. K1, K2, K3... Kn and id's of all sensor nodes and encrypted form of there ids with their corresponding key i.e. ($ID_A$) K1,($ID_B$)K2, ($ID_C$)K3,…… prior to deployment. The key is only available with the master node. Here sensor node A contains $ID_A$ and ($ID_A$)K1 and sensor node B contains $ID_B$ and ($ID_B$)K2 .
3. Sensor node A generates a random number i.e. Nonce A and computes the MAC using ($ID_A$)K1 and Nonce A .
4. Subsequently node A sends a message i.e. IDA||NonceA|| MAC {(IDA)K1|| Nonce A} to sensor node B.
5. Upon receiving the message, Node B tries to calculate the MAC for authentication. But as the key is only available with the master node so, sensor node B generates a random number i.e. Nonce B after getting the message from sensor node A and sends a message i.e. NonceA||NonceB|| $ID_B$ ||H{($ID_B$)K2}|| $ID_A$ to the master node to verify the identity of sensor node A.
6. Master node recognizes IDB and then authenticate by computing the hash over the {(IDB)K2 }which is with the master node. After authenticating node B from the message received from sensor node B and becomes sure about the validity of the message and confirm about the validity of sensor node A by sending (IDA)K1 to sensor node B.
7. Sensor node B then computes MAC {(IDA)K1|| Nonce A} and compares it with the MAC present in the message coming from sensor node A . If those two MACs match with each other then it becomes sure that Sensor node A is a valid node in the network and communication with it is safe.




**FigCaption1: Authentication Procedure**

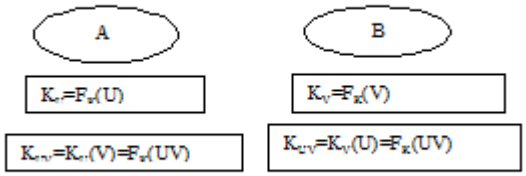
While node B communicating with the master node, if any adversary takes away all the communicated data ,it would be of no use as the (IDB)K2 of node B at node B had already been changed to (IDB)K2+ $N_{RANDA}$ , and similarly at the master node the change to the (IDB)K2 to(IDB)K2+ NonceA also takes place. So the data communicated between node B and master node if intercepted is of no use. Once master node authenticate node B the master node can directly send the Nonce B to node A .So that node A can update (IDA)K1 to (IDA)K1+NonceB and the same updation is also being carried out by the master node. Thus authentication between two sensor nodes is performed in the proposed scheme. Authentication is done solely by the master node.

### 3.2 Pair-Wise Key Establishment among two individual Sensor Node

In the pair wise techniques, generally a key used to establish a secure channel between a couple of nodes, where no one can use this channel to perform any type of communication without of the permission of the owner node of the key itself.

1. Let us consider there are two nodes having id U and V, want to communicate with each other. The node U computes its key $K_U$ as $K_U = F_K(U)$ .Here $F_K$ is a function and the node id U is an input to the function. Similarly the node V computes its key $K_V$ as $K_V = F_K(V)$.
2. Node U computes its pair wise key $K_{UV} = K_U (V) = F_K(UV)$ and node V computes its pair wise key $K_{UV} = K_V (U) = F_K (UV)$.
3. After the establishment of the pair wise key they start to transmit message to each other. Any message M is encrypted as $M= K_{UV}(M)$. A counter is set to keep records of transmitted message. With each transmission the

counter value is incremented at both the ends, so that the same message get encrypted differently each time which creates confusion for the attacker to decipher the key, as the key keeps changing with each transmission.




**FigCaption2: Pair wise key establishment mechanism**

4. After transmission of one message, the counter value is changed to some value C and the pair wise key between the sensor nodes becomes updated by that value i.e. $K_{UV} = F_K(UV+C)$. A message M can be encrypted as $M= K_{UV}(M) = F_K(UV+C)(M)$.

The main task of this scheme is to minimize the used memory for storing the different keys of the WSN, in addition to generating a strong and robust environment against any possible attack, and in case of any attack, the scheme should be flexible and fast in recovering.

### 3.3 Key management scheme for node revocation in the network

As the sensor nodes deployed in a network has the different limitations, hence in turn they get malicious or compromised and they hamper secure communication in the network by incorporating incorrect information or modifying data at other nodes in the network. This leads to the origination of an idea of assuming a private symmetric key called group key, which is used to transfer the messages within the group. When a compromised node is found, the current group key must be cancelled or revoked and a new key must be provided to all the group members except the compromised or malicious node. Hence it leads to the removal of the compromised Sensor node.

The protocol is efficient and scalable as it limits both communication and computation overhead during rekeying. This scheme follows an approach of node management (NM) that cancels the current group key and distributes a new group key to all the nodes except the compromised one and in turn forces the compromised nodes to leave the group.

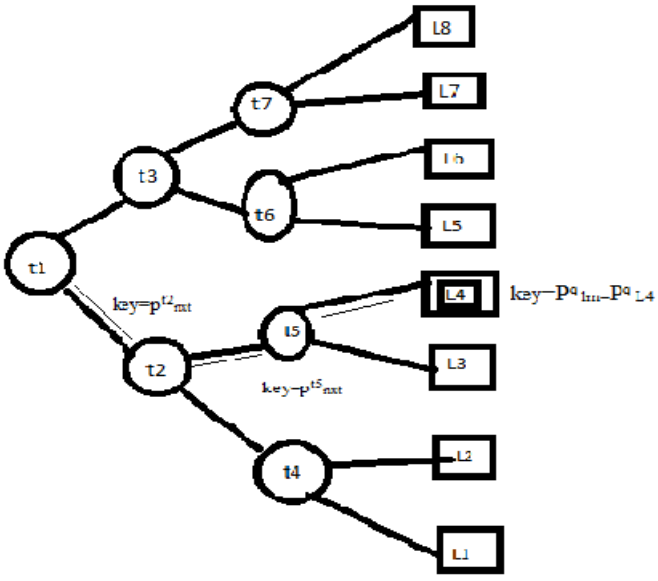
The node revocation procedure lies on the concept of group key set up phase, where a sender(G) wants to communicate with the receiver node(H) .In the key set up phase G creates a key-set, a sequence of values $< P_0 . . . P_i . . . P_N >$, where each key is linked to other ones by means of a hash function H. G randomly chooses the end-key $P_N$ and then iteratively applies the one-way function H for N times in order to obtain the start-key at the beginning of the chain. That is,

$$P_i = H(P_{i+1}), \quad (0 \leq i < N)$$

Each key in the set (except the start-key) is the hash preimage of the previous one in the direction of the start key. That is, the key $P_i$ is the hash preimage of $P_{i+1}$ .

The node revocation scheme is explained in the following steps.

1. Assume that each sensor node shares a secret symmetric key that shares with NM and is denoted by $P^q_{ln}$ i.e.; the key of the sensor node $l_n$.
2. To remove a malicious node $l_m$, NM maintains a stepwise structure of symmetric keys, called **the node-revocation tree.**
3. NM associates each internal node $t_a$ of the node revocation tree with a chain composed of $N_{ta}$ keys and each leaf node with the secret key of a group member. Initially, the current-key of the internal node $t_a$ , $P^{ta}_{cur,}$ corresponds to the start-key, $P^{ta}_0$.Thus, in the setup phase current keys contains only the start-keys.
4. A sensor node $l_n$ stores a set of current keys, key ring($l_n$) can be defined as, $KeyRing(ln) = \{P^{ta}_{cur} \mid ta \in Path(l_n)\}$, where Path($l_n$) be the set of internal nodes lying on the path from the root to the leaf associated with $l_n$. The key-set KeyRing($l_n$) is composed of the current-keys associated with the internal nodes in Path ($l_n$).
5. Every group member stores a copy of the current-key associated with the tree root, $P^{t1}_{cur}$, because it belongs to the key-set of every group member. Thus, this key acts as the group-key.
6. When a sensor node $l_m$ is compromised and NM has to renew them, but it isn't renewed for $l_m$.




**FigCaption3: Node revocation scheme**

7. For each internal node $t_a$ in Path($l_m$), and for each child of $t_a$, NM broadcasts a rekeying message structured as follows. Every message contains the next-key of $t_a$, $P^{ta}_{nxt}$. If the $t_a$ 's child is a leaf and it is not associated with $P^{q}_{lm}$, $P^{ta}_{nxt}$ is encrypted with the private-key of the corresponding group member. If the $t_a$'s child is an internal node (i,e; $t_c$) and is not included in Path($l_m$), $P^{ta}_{nxt}$ is encrypted with $P^{tc}_{cur}$. If $t_c$ is included in Path($l_m$), $P^{ta}_{nxt}$ is encrypted with $P^{tc}_{nxt}$.
8. After this procedure, for each internal node ta belonging to path $l_m$, $P^{ta}_{cur}$ is replaced by $P^{ta}_{nxt}$. Hence the leaf node with $P^{q}_{lm}$ is removed by the NM.

## 4. Conclusion

Security is the major concern in the wireless sensor networks. The communication of information between the sensor nodes gets tampered due to the malicious behaviour of the network. Hence in this paper, three schemes are discussed for secure routing between the sensor nodes using the concept of key management. The authentication scheme maintains an efficient and secure communication between two different nodes. The pair-wise key management scheme enforces security in terms of one private paired key. The third scheme explains one scheme for the problem of insecure behaviour of a compromised node, where the compromised node is revoked from the group to make the network more secure, efficient and scalable.

## References

[1] D.Estrin, R.Govindan, J.Heidemann, and S.Kumar, Next century challenges: scalable coordination in sensor networks, ACM MobiCom'99, Washington, USA, 1999, pp. 263–270.
[2] Laurent Eschenauer and Virgil D.Gligour.A Key Management Scheme for Distributed Sensor Networks. In proceedings of the 9th ACM conference on Computer and Communication Security, Pages 41-47, November 2002.
[3] H.chan, A.Perrig and D.Song, Random Key Predistribution Schemes for Sensor Networks. In IEEE Symposium on Security and Privacy, pages 197213, Berkeley, California, May 11-14, 2003.
[4] Pair wise Key Management Scheme Using Sub Key Pool for Wireless Sensor Networks by Jianmin Zhang, Yu Ding of Henan Institute of Engineering, China 2010 Second International Conference on Information Technology and Computer Science.
[5] A Promising Pair wise Key Establishment Scheme for Wireless Sensor Networks in Hostile Environments by WenqiYu of Henan Institute of Engineering, China.
[6] Stephen Anokye, Thabo Semong, Qiaoliang Li, Qiang Hu ,"Group wise Pair wise Scheme for Wireless Sensor Networks," niss, pp.695-699, 2009 International Conference on New Trends in Information and Service Science, 2009
[7]A Pair wise Key Establishment for Wireless Sensor Networks Found in: 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing By Hung-Min Sun, Yue-Hsun Lin ,Cheng-Ta Yang , Mu-En Wu
[8] A new pair wise key establishment scheme for sensor networks By Sujun Li; Siqing Yang; Suying Zhu; Fuqiang Yan; Dept. of Comput. Sci.&Technol., Hunan Inst. Of Humanities, Loudi, China in.Computer Application and System Modeling (ICCASM), 2010 International
Conference.
[9] D. Liu and P.Ning, "Establishing pair wise keys in distributed sensor networks," in Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS),Washington, DC, USA, October 27-31 2003, pp. 52–61.
[10] W. Du, J. Deng, Y.S.Han , and P.K.Varshney ,"A pair wise key predistribution scheme for wireless sensor networks," in Proceedings of the 10th ACM Conference on Computer and
Communications Security(CCS),Washington, DC,USA ,October 27-31 2003, pp. 42–51.
[11]A robust group-based key management scheme for wireless sensor networks By Zhen Yu Yong Guan Dept. of Electr. & Comput. Eng., Iowa State Univ., Ames, IA, USA Wireless Communications and Networking Conference, 2005 IEEE.
[12] Sub-Grid based Key Vector Assignment : A Key Pre Distribution Scheme For Distributed Sensor Networks(2006) By R. Kalindi , R. Kannan, S. S. Iyengar and A. Durresi of
Louisiana State University, USA.

**First Author Soumyashree Sahoo** continuing her MTech. Degree in Computer Science and Engineering from BPUT. She is currently a lecturer in Department of Computer Science and Engineering of Hi-tech Institute of Technology, Khurda. Her research interest includes Wireless sensor network, Cryptology and Network security.

**Second Author Pradipta Kumar Mishra** received his MTech. Degree in Computer Science and Engineering from IIIT-BH. He is currently an Assistant Professor in Department of Computer Science and Engineering of Hi-tech Institute of Technology, Bhubaneswar .His research interest include Ad-Hoc network, Wireless sensor network and Network security.

**Third Author** Prof.(Dr.) Rabi Narayan Satpathy received his PhD. Degree in Computational Mathematics and Computer Science & Engineering from Utkal University ,Odisha and cosmopolitan University and PDF From NIT,Rourkrla,Odisha & D.Sc. in computational Fluid Dynamics from FM University . He is currently the Principal in Hi-tech Institute of Technology, Bhubaneswar. His research interest include Soft Computing , Ad-Hoc network, Wireless sensor network and Network security.

# A New Spectral Based Characterization of Electrocardiogram Signals in Sudden Cardiac Death

**Hedi Khammari**

**College of Computers and Info. Tech., Taif University, Saudi Arabia**

**Abstract**

A new method of characterization of patient Electrocardiograms (ECG) suffering from Sudden Cardiac Death (SCD) permits to confirm the role of lower order higher harmonics when approaching the moment of SCD instant. Besides, several peak amplitudes are detected during the normal heart beat oscillations preceding such cardiac arrest. Abrupt qualitative changes often encountered in the behavior of nonlinear dynamical systems may be regarded as analogous to the main features of the changing dynamics of the heartbeat. Indeed, under certain circumstances, the heart rhythm oscillations may undergo an abrupt change in frequencies and amplitudes of spectral lines. A sliding window FFT performed on ECGs of two patients suffering from Sudden Cardiac Death, leads to put into evidence the reorganization of the higher harmonics involving changes in their ranks and their amplitudes.

***Keywords:*** *Electrocardiogram (ECG), Spectral Analysis, Sudden Cardiac Death (SCD), sliding window (FFT), Higher Harmonics.*

## 1. Introduction

The human circadian rhythms, in normal or abnormal situations or in healthy and diseased cases, present complex oscillations that undergo several changes under either external or internal excitation: sleep/awake, happiness/sadness, light/darkness. The heart oscillatory behavior can be examined in frequency domain, in order to reveal what are the harmonics that have major effect in such or such state. Several research studies based on spectral analysis of ECGs have revealed important results [1], [5], [7], [8], [11]. In former studies, it was shown that the harmonic predominance can characterize certain bifurcation structures in a nonlinear electric circuit [2]. Similarly, ECG is a nonlinear oscillation that may be characterized by the spectral line ranks changes.

The obtained results in nonlinear system theory, offer new descriptive and perspective insights into physiological systems that may more reflect complex behaviors such as bifurcation and chaos. Nonlinear dynamic behavior seems to occur in the heartbeat time-series data. Thus, the analysis of some recordings of ECGs by means of nonlinear theory approaches, may lead to highlight the existence of nonlinear phenomena. Methods from nonlinear dynamics have revealed new insights into heart rate variability (HRV) changes under various physiological and pathological conditions leading to provide further prognostic [3], [4], [12]. Sudden Cardiac Death refers to natural cardiac death due to loss of the body consciousness caused by heart's electrical system malfunctions. The most common cause of SCD is a heart rhythm disorder or arrhythmia called ventricular fibrillation (VF). The cardiac arrest occurs for a very short time which is preceded and followed by normal ECG [10]. Patients at high risk of sudden cardiac death show evidence of nonlinear heart rate dynamics, including abrupt spectral changes and sustained low frequency (.01-.04 Hz) oscillations in heart rate. Characteristically, frequency domain methods are used for quantifying the autonomous nervous system control as well as to guess sudden cardiac death [6]. According to [16], despite the remarkable progress in the field of prediction and prevention of SCD, there is always a need to develop new appropriate methods for reliable testing of the clinical utility of known risk variables for the prediction of arrhythmic death.

The main features of a normal ECG rhythm of patient suffering from Sudden Cardiac Death are mainly the lower spectral energy and the low frequency range of first lobe [10]. Power spectral methods analysis of HRV is a useful noninvasive tool that permits to categorize cardiac patients according to risk of SCD [9].

Some Holter variables can predict the occurrence of SCD among the patients with slightly reduced or preserved left ventricular (LV) function [15]. Other tools of prediction and prevention of SCD have been reported in [14], [17], such as T-wave alternans (TWA) and Modified Moving Average (MMA).Such tools were used to identify individuals at heightened risk for SCD. Spectral and MMA-TWA would both predict arrhythmia-free survival [8]. It was stated that risk of SCD rises when a cutoff point

of 46 μV TWA is exceeded [14]. The application of FFT on QRS complex used to provide information from normal ECG signals was sufficiently performed on small portions (4-5 minutes) of any patient ECG to detect possibility of SCD [10]. The remainder of this paper is split into four sections; section 2 contains the basic reminders and the methodology. An experimental setup is given in section 3. Section 4 presents the main obtained results. Finally, section 5 is devoted to interpret such results.

## 2. Methodology

In recent literature, approaches to understanding and intervening in the cardiovascular system are being developed using the new methods from nonlinear system theory. Spectral analysis of ECG signals is a preliminary study aiming to classify certain particular physiological signal through their spectral composition and the associated ranks of higher harmonics so that, such analysis may become an important tool to reveal particular healthy or diseased situations. It is well known that a typical heart beat has the extremes P, Q, R, S and T, the beat-beat interval is the time between two consecutive R-peaks. This can be considered as the period of heart oscillation.

This study aims to characterize Electrocardiogram (ECG) signals, for patient suffering from SCD, through the reorganization of the higher harmonics. Approaching the SCD moment, the spectra of ECGs include higher harmonics which undergo changes of their ranks and their amplitudes in a certain classification defined as follows: Let r be the place occupied by an order-p harmonic spectral line from an ordering based on the amplitudes of spectral lines in descending order. So if r=2, the order-p harmonic is said to be simple predominant and if r=1, it is a full predominant one.

The spectral composition of an ECG signal depends on its clinically relevant parameters such as time interval between waves, duration of each wave or composite waveforms, peak amplitudes. ECG exhibits clinical information from generation and propagation of electric signal in the hearts. Among our objectives, we attempt to localize abnormality in frequency domain as well as in time domain. Aiming to analyze Electrocardiogram (ECG) signals given as discrete time signals, one can investigate how the frequency content varies with time. Defining a sliding window FFT algorithm, we proceed with two different ways: separated windows or overlapped windows. We choose a short window size (of 2024 points) and slide this window across the 15000 data samples while taking the 2024 point FFT and shifting the window by more than 2024 samples for separated windows or by less than 2024 samples for overlapped windows. After doing so, continue this operation until the end of the 15000th sample. The 2024 points FFT provides one spectrum, represented as a 1D vector that describes the whole data segment.

The result from applying the sliding window FFT is in fact a 2D image of dimension frequency versus time that sketches the spectrum for different time intervals. This image is known as a spectrogram.

## 3. Experimental Setup

The current study included only two subjects and relied on spectral analysis of ECG signals. From Sudden Cardiac Death Holter Database (Physionet)[13], the selected waveform records 30 (male 43 years old) and 47 (male 34 years) are of two patients suffering from a cardiac arrhythmia. Such patients experienced sudden cardiac death during the recordings. The spectral analysis of their Electrocardiograms leads to study in depth the main frequency features of the heart rhythm signals.

Aiming to characterize the heart beat oscillations before and during the sudden change, we sketch the continuous change of the spectral composition corresponding to a sliding window taken on the temporal data of the ECG. Every spectrum is associated to a certain classification based on decreasing amplitudes of the spectral lines, including the constant Fourier coefficient, the fundamental and the higher harmonics. In a first attempt, we perform a spectral analysis of an one minute length recording for patient 30, within which the sudden cardiac death happens. Such interval is split into separated or slightly overlapped windows. Such analysis devotes to highlight the spectral reorganization prior the SCD, is then applied to strongly overlapping intervals for both patients 30 and 47.

In a second attempt the spectral analysis is applied to a 7-minute interval preceding the SCD moment in order to observe the earlier changes in the ECGs spectra for the considered patients.

## 4. Results

*Spectral Reorganization*

Using the sliding window FFT, the ECG changes in temporal domain are accompanied by spectral changes. Some properties are basically deduced from spectral rather than from temporal signal representation. Following the ECG dynamics through spectral analysis leads to underline the spectral reorganization going with the transition from healthy to diseased cardiac rhythm state or

vice versa. In the following we summarize some obtained results corresponding to the patient 30, this study focuses on one minute interval including SCD after nearly 40 seconds. We realize that approaching the SCD instant the spectra undergoes several changes illustrated by the higher harmonics descending order classification defined above. Each temporal window is characterized by the so called HH ordering as follows:

- Interval A [0, 2.5*s*]: 1, 2, 6, 5, 9: the fundamental harmonic is in the first rank the other higher harmonics have smaller amplitudes, see figure 1.
- Interval B [2, 4.5*s*]: 1, 2, 5, 8, 6, 0: the $6^{th}$ higher harmonic is moved from the $3^{rd}$ to the $5^{th}$ rank and the $8^{th}$ HH gains the $4^{th}$ rank of the classification, see figure 2.
- Interval C [11, 14*s*]: 0, 1, 2, 3, 5: the dc Fourier coefficient becomes larger in amplitude than all the rest of harmonics, and occupies the $1^{st}$ rank: permutation of HHs 3 and 5.
- Interval D [14, 16.5*s*]: 0, 1, 2, 3, 6, 8: no changes in the 4 first ranks.
- Interval E [16, 18.5*s*]: 1, 8, 5, 6, 13: several changes: the higher harmonics of orders 2 and 3 are shifted from the $2^{nd}$ and the $3^{rd}$ rank respectively.
- Interval F [18, 21*s*]: 5, 3, 8, 2, 13 : The fundamental harmonic is no more among the first ranks, and the 5*th* higher harmonic is predominant.
- Interval G [21.5, 24.5*s*]: 0, 8, 5, 3, 13, ..: permutation between HH 8 and 5, 3.
- Interval H [30, 32.5*s*]: 0, 7, 3, 5, 8, 1, ..: the fundamental harmonic is in the $5^{th}$ rank.
- Interval I [40, 42.5*s*]: 1, 8, 13, 5, 0, 3, ..: the fundamental harmonic is in the $1^{st}$ rank.
- Interval J [42, 44.5*s*]: 1, 0, 2, 7, 8, 3, ..: closer to the beginning of the sudden heart death, the lower order HH gain the first ranks.
- Interval K [44, 46.5*s*]: 0, 1, 5, 3, 2, ...: more odd harmonics among the first ranks.
- Interval L [46, 48.5*s*]: 0, 1, 2, ..: very important amplitude of the dc Fourier coefficient, then the fundamental harmonic, then the $2^{nd}$ HH with significant amplitude and the remainders HH with vanishing amplitudes.
- Interval M [48, 50.5*s*]: 1, 2, 0, 6, 3, ...: permutation of 0 and HH 1,2.
- Interval N [50, 52.5*s*]: 0, 4, 2, 1, 6, 8, ..: besides the dc Fourier coefficient, even order harmonics becoming important in amplitudes.
- Interval O [52, 54.5*s*]: 2, 4, 0, 6, 8, 5, 1, .. : Higher harmonics of even orders become in the first ranks and the odd order ones are shifted far in the classification.
- Interval P [57, 60*s*]: 4, 2, 6, 8, 19, 12, .. :More even higher harmonics in the first ranks.

The transition between two overlapping intervals involves minor changes compared to those that appear between two disjoint intervals. Such changes may be even less if the intervals are strongly overlapped.

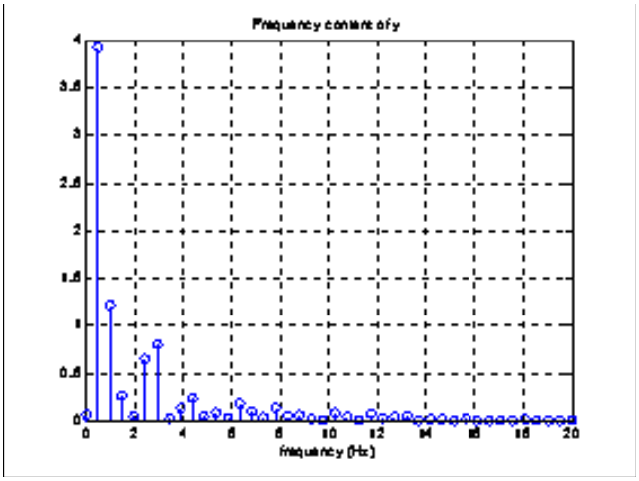



Fig. 1 Spectrum for Range A.

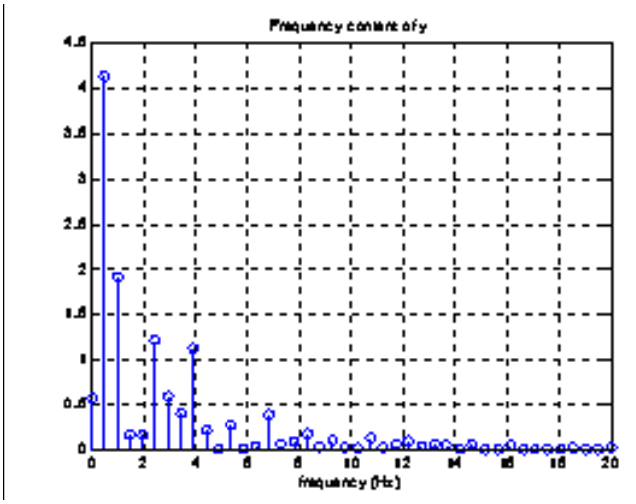



Fig. 2 Spectrum for Range B.

*Overlapped Sliding Windows*

Aiming to carry out an efficient spectral prediction tool of SCD, we defined common width windows of 2024 samples (nearly 8 s) and a sliding time step of 256 samples (almost 2 s). Over a 1-minute interval waveform 56 windows are generated. We attempt to analyze the few minutes preceding the SCD instant. We compute the corresponding spectrum of each window of a time series. Emphasizing on the most relevant harmonics over each 1-minute ECG spectrum, we define for each spectrum four sets Si, i=1,4 including either dc Fourier coefficient,

fundamental harmonic and higher harmonics occupying the *ith* rank in the amplitudes classification in descending order.

*Windows including SCD instant*

The 1-minute waveforms given in Figures 3 and 8 for patients 30 and 47 respectively, including the instant of SCD, are analyzed with mean of the technique described above. Since we are interested in the four higher harmonics occupying the first ranks, the HH predominance sets S1, S2, S3 and S4 of patient 30 are obtained from Figures 4, 5, 6 and 7 respectively. S1={0, 1, 2, 3,4, 46, 52, 55, 56, 57, 73}, S2={0, 1, 2, 3, 4, 43, 51, 52, 54, 55, 56, 73}, S3={0, 1, 2, 3, 4, 5, 44, 46, 47, 48, 52, 56, 57, 60} and S4={0, 1, 2, 3, 4, 5, 9, 42, 51, 52, 55, 56, 60, 72}. For the patient 47 such sets can be derived from Figures 9,10,11 and 12 : S1={0, 1, 2, 3, 4, 22, 24, 29, 30, 37}, S2={0, 2, 3, 4, 21, 22, 23, 24, 26, 28, 29, 31, 33, 34, 35}, S3={0, 1, 2, 3, 4, 9, 10, 14, 16, 21, 22, 23, 24, 25, 26, 28, 30, 31, 33, 34, 38} and S4={0, 1, 2, 3, 4, 9, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 30, 31, 32, 34, 37}. Looking to the evolution of each higher harmonic amplitude separately; one can underline the higher harmonic rank changes in the sliding window FFT going toward the SCD instant and point out the existence of a peak amplitude in each 1- minute interval associated to the fundamental harmonic for record 30 and to the third order higher harmonic for record 47.The amplitudes variations of the spectral lines along a 1-minute interval just before SCD are given in figures 13 and 14 for the patients 30 and 47 respectively. After SCD the 4 first ranked higher harmonics which are mainly the low order or constant Fourier transform components the (0, 1, 2, 3).

*One minute before the SCD*

The spectral analysis of the 1-minute interval preceding the SCD for both of the patients 30 and 47 leads For the patient 30: S1={0, 10, 20, 21, 30, 31}, S2={0, 10, 20, 21, 30, 31}, S3={0, 10, 11, 20, 21, 30, 31,32}, S4={0, 10,11, 20, 21, 30, 31,32,51,53}. During such minute of record 30, the main HH occupying the first ranks are besides the dc component, the HH 10 (1.22 Hz), 20 (2.44 Hz), 30 (3.66 Hz), this may be a spectral characterization of the patient 30 ECG in a normal situation preceding the SCD. For the patient 47, the 4 sets of predominant harmonics during the 1-minute interval foregoing the SCD are: S1={0, 3,4}, S2={0, 2, 3, 4, 11, 14}, S3={0, 1,2,3, 11, 15}, S4={0, 1, 2, 3, 5, 9, 11, 14, 15}. It is obvious to underline the fact that the lower order HH are already predominant, seemingly the HH which typify the normal state are shifted to higher ranks. Certain distortions of the ECG may affect the symmetry of the signal and leads to an important Fourier constant component. The largest higher harmonic exhibit peak amplitudes in both cases, in tables 1 and 2 are summarizing the main features of the 7-minutes intervals preceding SCD of patient 30 and patient 47 respectively.

*Seven minutes interval prior to the SCD*

The spectral behavior of a seven minutes interval preceding SCD is analyzed for the two patients 30 and 47. The orders of the higher harmonics occupying the first ranks of the sliding windows spectra are given in figures 15 and 16. Regarding the amplitudes, the most relevant harmonics are those in the low frequency range 0-1.22 Hz (figures 17 and 18), whereas those in the frequency range 1.22-12.2 Hz (figures 19 and 20) are least significant.

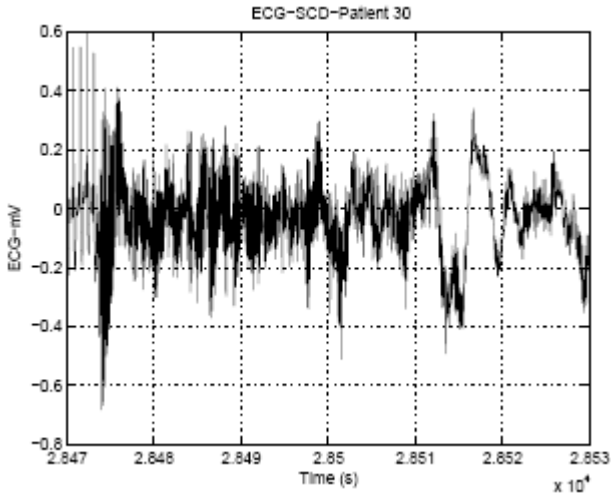



Fig. 3 ECG Waveform for Patient 30.

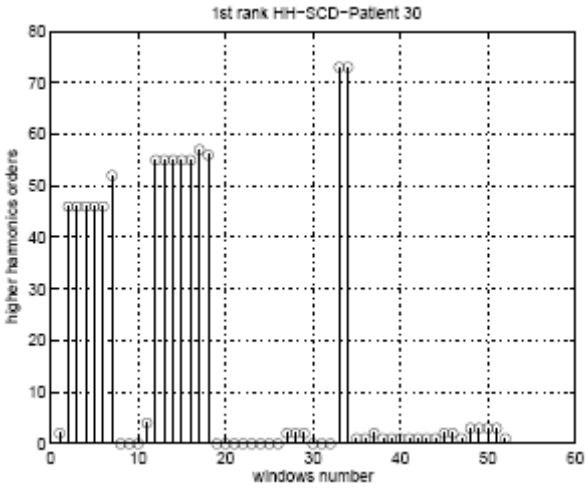



Fig. 4 HH in the first rank of the amplitude classification.

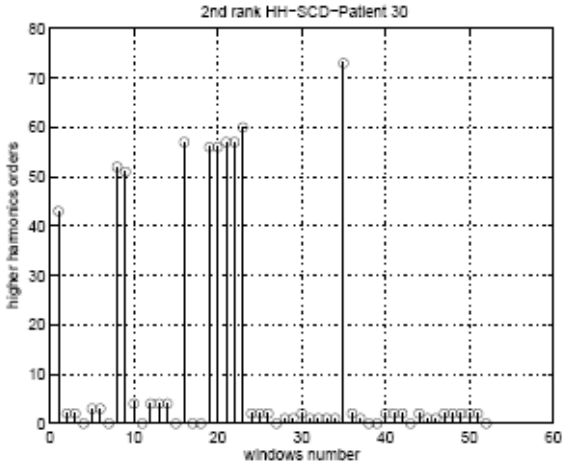



Fig. 5 HH in the second rank of the amplitude classification.

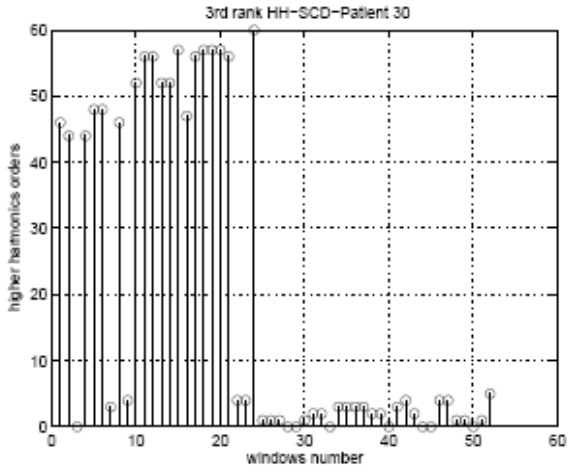



Fig. 6 HH in the third rank of the amplitude classification.

## 5. Discussion

The higher harmonic predominance is used in this study to characterize the ECGs dynamics of patient suffering from SCD. In [3] it was stated that patients at high risk of sudden cardiac death show evidence of nonlinear heart rate dynamics, including abrupt spectral changes and sustained low frequency (.01-.04 Hz) oscillations in heart rate rhythm. Approaching the SCD moment, the common phenomena observed in both records 30 and 47 is in the first hand, the higher harmonics laying in the low frequency range become progressively among the first ranks in the decreasing amplitude classification. In the second hand, the occurrence of peaks of HH amplitudes during the few minutes preceding the SCD. Therefore, Further information of primary interest emphasizing on the role of higher harmonics complement prediction tools focusing on amplitudes to contribute to design new SCD predictors.

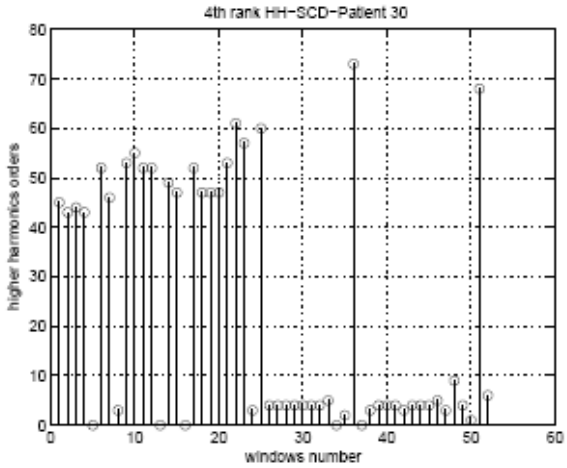



Fig. 7 HH in the fourth rank of the amplitude classification.

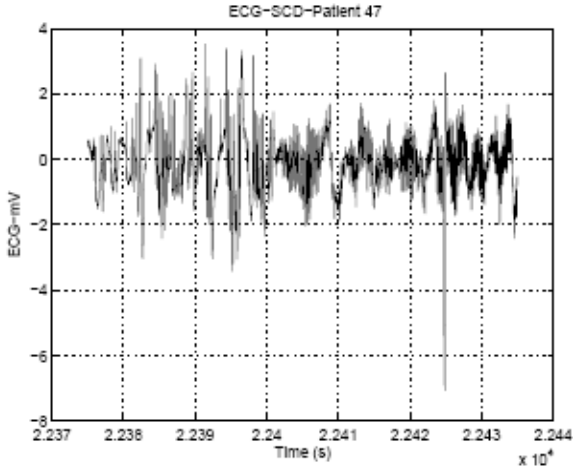



Fig. 8 1-minute ECG waveform for Patient 47.

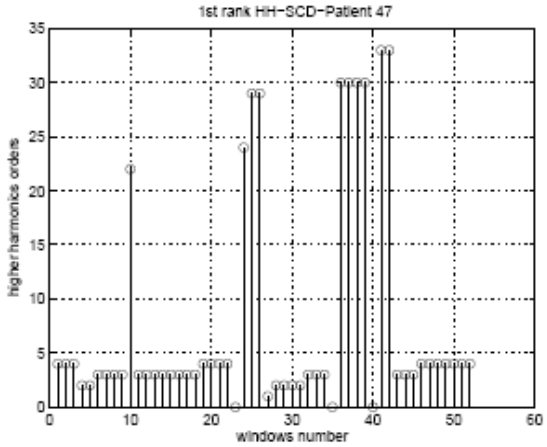



Fig. 9 HH in the first rank of the amplitude classification.

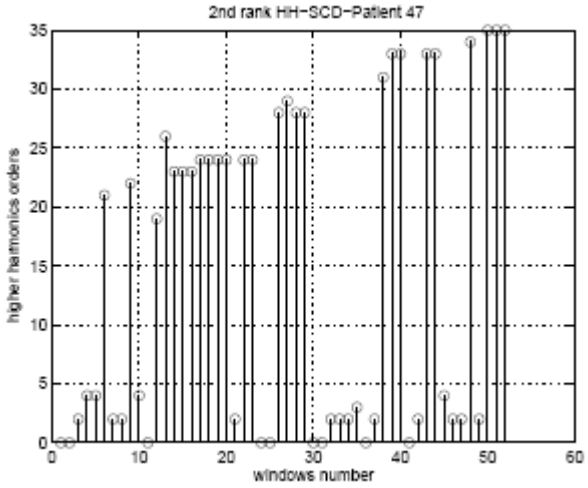



Fig. 10 HH in the first rank of the amplitude classification.

In tables 1 and 2 are given the peak amplitudes over a 1-minute interval and the mode harmonic value which is the harmonic that is frequently in the first rank.

Table 1: Largest HH in ECG Spectra: Patient 30.

| Time | Largest HH in 1st rank | Peaks | Mode |
|---|---|---|---|
| SCD | 0 1 2 3 4 46 52 55 56 57 73 | 1 | 0 1 |
| -1mn | 0 10 20 21 30 31 | 21 10 | 20 10 |
| -2mn | 0 1 20 21 29 30 31 | 0 | 0 20 |
| -3mn | 0 1 2 4 26 | 0 | 0 |
| -4mn | 0 1 | 0 | 1 0 |
| -5mn | 0 1 9 13 14 18 | 0 | 13 0 |
| -6mn | 0 9 13 14 27 | 0 | 0 13 |
| -7mn | 0 13 14 | 0 | 0 14 |

Table 2: Largest HH in ECG Spectra: Patient 47.

| Time | Largest HH in 1st rank | Peaks | Mode |
|---|---|---|---|
| SCD | 0 1 2 3 4 22 24 29 30 37 | 1 | 3 4 |
| -1mn | 0 3 4 | 3 4 | 4 |
| -2mn | 0 1 3 4 | 0 1 | 4 0 4 |
| -3mn | 0 2 3 4 7 11 13 15 | 11 15 | 0 4 |
| -4mn | 0 1 2 4 5 6 7 8 | 1 2 | 0 1 |
| -5mn | 0 1 2 3 4 5 6 7 | 0 | 0 |
| -6mn | 0 1 2 3 4 6 7 | 1 | 0 1 |
| -7mn | 0 1 3 4 6 7 | 0 | 0 |

The dc Fourier constant is omnipresent among the four largest amplitudes of the descending order classification, besides many peaks correspond to such dc component in the case of patient 30. Focusing on the 1-minute intervals just before the SCD one can observe certain similarities between the two studied cases. The Higher harmonics in first ranks (0, 3, 4) for record 47 and (0 10, 20, 21, 30, 31) for record 30 have relatively closer higher amplitudes. The spectra of first intervals just before the SCD of patient 30 exhibit the abrupt occurrence of higher frequencies 46 52 55 56 57 (5.6 - 6.95 Hz). For the patient 47 the higher frequencies which appear mainly in the second rank (1-5 Hz), during the 7 minutes foregoing the SCD the spectra show the predominance of lower order higher harmonics (0.122-0.6 Hz).

The major changes may affect the amplitude, the ranks and the orders or frequencies of higher harmonic forming the spectra of ECG. The peak amplitudes which may coincide with premature ventricular contractions is the common features revealed from the spectra of 7-minute interval preceding SCD for both patients 30 and 47.Such peak amplitudes correspond mainly to the lower order higher
harmonics.

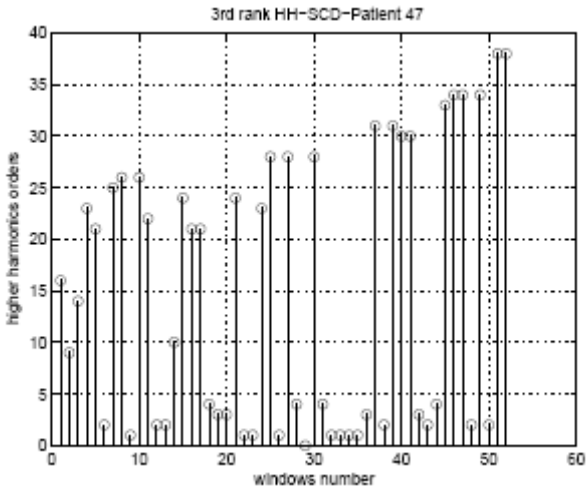



Fig. 11 HH in the first rank of the amplitude classification.

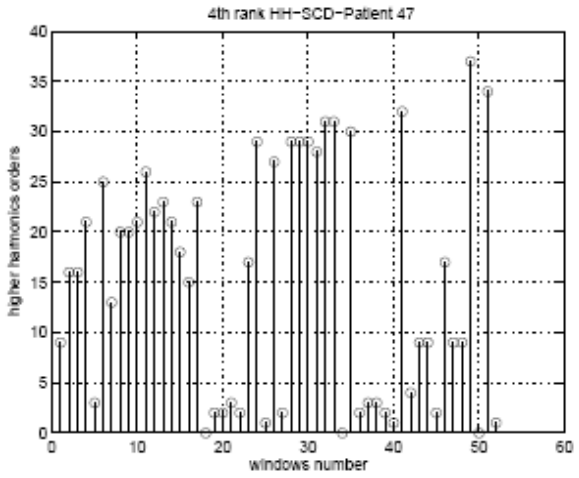



Fig. 12 HH in the fourth rank of the amplitude classification.

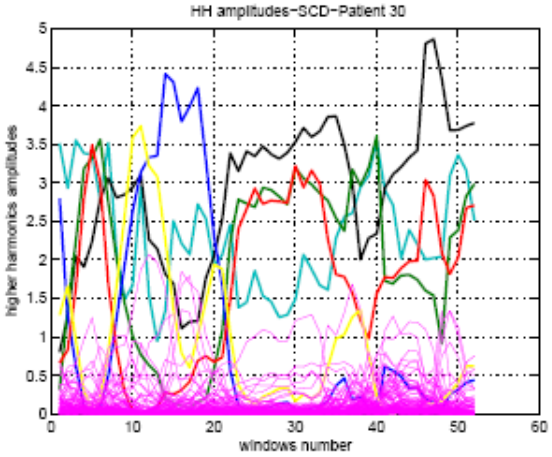



Fig. 13 HH amplitude variation along a 1-minute interval just before SCD (Patient 30).

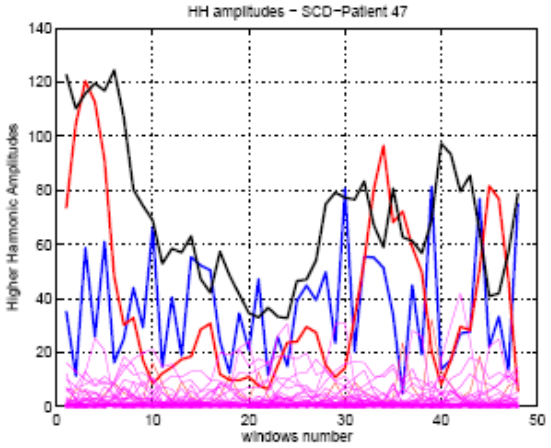



Fig. 14 HH amplitude variation along a 1-minute interval just before SCD (Patient 47).

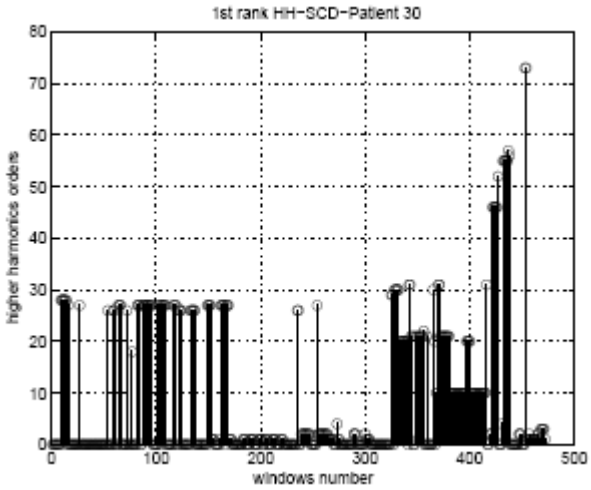



Fig. 15 HH in the first rank of the amplitude classification along a 7-minutes interval just before SCD (Patient 30).

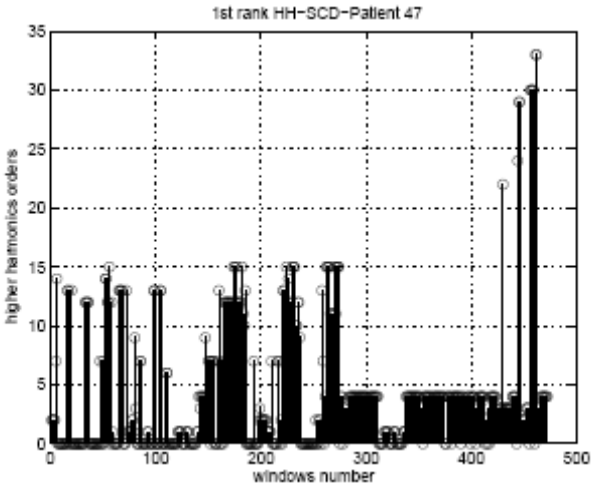



Fig. 16 HH in the first rank of the amplitude classification along a 7-minutes interval just before SCD (Patient 47).

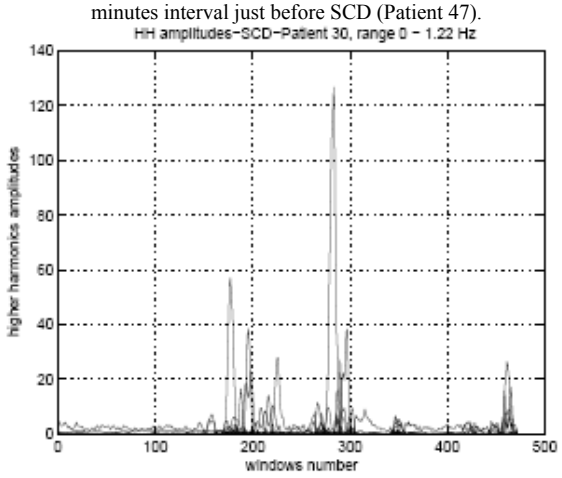



Fig. 17 Evolution of the lower order HH (0 – 1.22 Hz) during an interval of 7-minutes forgoing SCD (Patient 30).

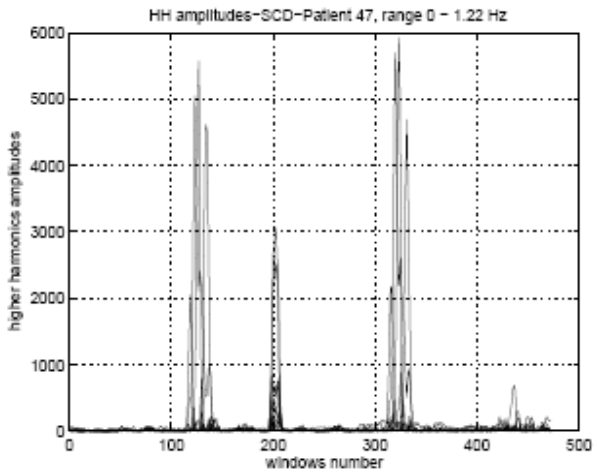



Fig. 18 Evolution of the lower order HH (0 – 1.22 Hz) during an interval of 7-minutes forgoing SCD (Patient 47).

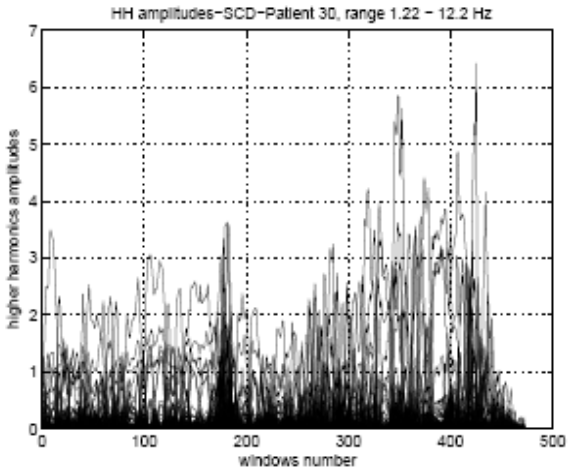



Fig. 19 Evolution of the higher order HH (0 – 1.22 Hz) during an interval of 7-minutes forgoing SCD (Patient 30).

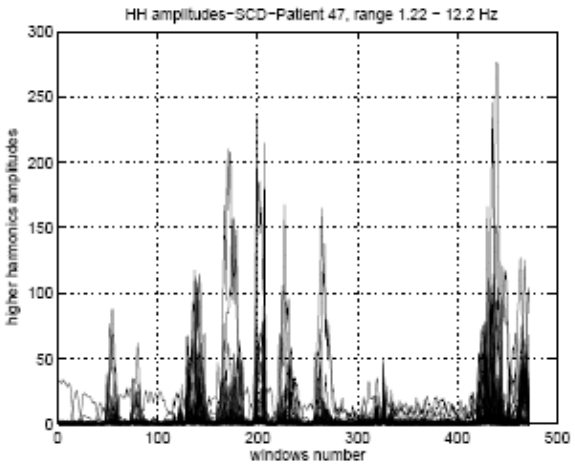



Fig. 20 Evolution of the lower order HH (0 – 1.22 Hz) during an interval of 7-minutes forgoing SCD (Patient 47).

## 6. Conclusion and future work

To the best of our knowledge, this study is the first to characterize the time preceding SCD through the higher harmonic interaction based on a descending order classification of their amplitudes. Spectral analysis of ECG signal of a patient who experienced SCD permits to put evidence into a complex spectral line reorganization that occurs before the SCD instant. Approaching the sudden cardiac death, the spectral composition of ECG signal undergoes quantitative changes which affect the amplitudes and the ranks of the spectral lines. The progressive occurrence of lower order harmonics and the peak amplitudes of harmonics from the same range in the few minutes preceding SCD, are common features observed for two different patients suffering from Sudden Cardiac Death. Further work might involve spectral analysis based on several human subjects with underlying Sudden Cardiac Death in order to broaden the scope of harmonic classification in descending order amplitudes to heart rate variability (HRV) regarding the R-R intervals.

### Acknowledgments

We thank Physionet.org for providing sudden cardiac death ECG data.

## References

[1] Sesay M. et al','Spectral analysis of the ECG R-R interval permits early detection of vagal responses to neurosurgical stimuli', Annales franaises d'anesthsie et de ranimation Vol. 22,N5, pp.421-424, 2003.

[2] Khammari H., Mira C. and Carcasses,J.P., 'Behavior of harmonics generated by a Duffing type equation with a nonlinear damping'. Part I: International Journal

of Bifurcation and Chaos, Vol. 15, N10, pp. 3181-3221, 2005.

[3] Goldberger A.L., Rigney D.R., Mietus J., Antman E. and Greenwald S., 'Nonlinear dynamics in sudden cardiac death syndrome: heart rate oscillations and bifurcations', Experientia 1988; 44:983-987.

[4] Goldberger A.L., Rigney D.R. On the non-linear motions of the heart: fractals, chaos and cardiac dynamics' In: Goldbeter A, ed. Cell to Cell Signaling: From Experiments to Theoretical Models. San Diego: Academic Press, 1989, pp. 541-550.

[5] Stoner R., 'T wave Spectral Analysis As A Marker For An Increased Risk Of Sudden Cardiac Death Using an in vivo Canine Model', Ohio State University, Department of Electrical and Computer Engineering Honors Thesis 2008.

[6] Claria F.et al., 'Time-Frequency Representation of the HRV:A Tool to Characterize Sudden Cardiac Death in Hypertrophy Cardiomyopathy Patients',in Proc. of the 22nd Annual EMBS International Conference, July 23-28,2000, Chicago IL.

[7] Kamada T., Miyake S.,Kumashiro M., Monou H., and Inoue K.,'Power Spectral Analysis of Heart Rate Variability in Type As and Type Bs during Mental Workload', Psychosomatic Medicine vol.54, pp.462-470, 1992.

[8] Cox V. et al.,'Predicting Arrhythmia-Free Survival Using Spectral and Modified-Moving Average Analyses of T-Wave Alternans', Pace, Vol. 30 , March 2007.

[9] Myers G.A. et al.,'Power Spectral Analysis of Heart Rate Variability in Sudden Cardiac Death: Comparison to Other methods' IEEE Transactions on Biomedical Engineering, Vol. BME- 33, No. 12, December 1986.

[10] Rashed U., and Mirza M.J., 'Identification of Sudden Cardiac Death Using Spectral Domain Analysis of Electrocardiogram (ECG)'International Conference on Emerging Technologies IEEE-ICET'08 Rawalpindi, 18-19 October 2008.

[11] Khadra L. , AlFahoum A.S., and Binajjaj S., 'A Quantitative Analysis Approach for Cardiac Arrhythmia Classification Using Higher Order Spectral Techniques' , IEEE Transactions on Biomedical Engineering, Vol.52,No.11, 2005.

[12] Voss A., Shultz S., Shroeder R., Baumert M. and Caminal P., 'Methods derived from nonlinear dynamics for analyzing heart rate variability' Phil. Trans. R. Soc. A ,vol. 367, no.1887, pp. 277-296, 2009.

[13] Heart Rhythm Society. www.HRSpatients.org, 'Sudden Cardiac Death Holter Database' http://www.physionet.org/pn3/sddb/

[14] Richard L. Verrier R.L., Kumar K.,and Nearing B.D., 'Basis for Sudden Cardiac Death Prediction by T-Wave Alternans from an Integrative Physiology Perspective' Heart Rhythm. 2009 March ; 6(3):pp.416-422.

[15] Makikallio T.H.et al,'Prediction of sudden cardiac death after acute myocardial infarction: role of Holter monitoring in the modern treatment era' European Heart Journal (2005) 26,762- 769.

[16] Huikuri H.V. et al,'Prediction of Sudden Cardiac Death: Appraisal of the Studies and Methods Assessing the Risk of Sudden Arrhythmic Death' Circulation 2003;108;110-115.

[17] Wood N.B.,'The Prediction of a Potentially Fatal Cardiac Event in the Next 2 to 24 Hours and The Prediction of a Myocardial Infarction Related Death or Sudden Death' Computers in Cardiology 2001;28:509-512.

Hedi KHAMMARI, PhD. He was born in Kairouan, Tunisia in 1963. He received the engineer diploma and the Master degree from National Engineering School of Tunis in 1988 and 1990 respectively. He received PhD in Electrical Engineering in 1999. He is currently Associate Professor at Taif University, Saudi Arabia. His research interests are mainly in the area of nonlinear dynamics and the application of chaos theory in different fields namely communication, electric systems and bioinformatics.

# Improve Data Warehouse Performance by Preprocessing and Avoidance of Complex Resource Intensive Calculations

**Muhammad Saqib[1], Muhammad Arshad[2], Mumtaz Ali[3], Nafees Ur Rehman[4], Zahid Ullah[5]**

**[1,2] City University of Science & Information Technology,
Peshawar, KPK, Pakistan**

**[3, 4, 5] Institute of Management Sciences,
Peshawar, KPK, Pakistan**

**Abstract**

A Data Warehouse is a computer system designed for archiving and analyzing an organization's historical data, such as sales, customers, products, salaries, or other information from day-to-day operations OLTP. Normally, an organization summarizes and copies information from its operational systems to the data warehouse on a regular schedule, such as daily, weekly, monthly, quarterly or annually; after that, management can perform complex queries and analysis OLAP on the information without slowing down the operational systems. Materialized views can be one best option in this regard and can be used in a number of ways. It can be used in distributed databases for replication and can also be used for efficient provision of data to a query through query re-writing. The process of data provision to queries can further be expedited if dependent child views are created on an already existing materialized view. Furthermore, these child-views are automatically created upon the creation of the base materialized view with some restrictions. This results in less-user dependent activity of creation of materialized views based on some parameters. These parameters are the number of child-materialized views and the type of the data a view contain. In this paper, a balanced approach is suggested to create sub-materialized views to answer user queries without consulting the fact table or parent materialized view that results in avoidance of resource intensive calculations and joining of multiple tables.

***Keywords:*** *Materialized View, Aggregation Plan, OLTP, OLAP.*

## 1. Introduction

Most of the modern enterprises and organizations rely on knowledge-based management systems. In such kind of systems, knowledge is gained from data analysis. Nowadays, knowledge-based management systems include data warehouses as their core components. The purpose of building a data warehouse is twofold. Firstly, to integrate multiple heterogeneous, autonomous, and distributed data sources within an enterprise. Secondly, to provide a platform for advanced, complex, and efficient data analysis. Data integrated in a data warehouse are analyzed by the so-called On-Line Analytical Processing (OLAP) applications designed among others for discovering trends, patterns of behavior, and anomalies as well as for finding dependencies between data. Massive amounts of integrated data and the complexity of integrated data that more and more often come from WEB-based, XML-based, spatio-temporal, object, and multimedia systems, make data integration and processing challenging.[1]

Information and knowledge is one of the most valuable assets of an organisation and when used properly can assist in intelligent decision making that can significantly improve the functioning of an organisation. Data Warehousing is a recent technology that allows information to be easily and efficiently accessed for decision-making activities by collecting data from many operational, legacy and possibly heterogeneous data

sources. On-Line Analytical Processing (OLAP) tools are well-suited for complex data analysis, such as multi-dimensional data analysis, and to assist in decision support activities while data mining tools take the process one step further and actively search the data for patterns and hidden knowledge in the data held in the warehouse. In our common practice, many organisations are building, and or planning to develop, data warehouses for their operational and decision support needs.[2]

Materialized views have been found to be very effective in speeding up query, as well as update processing, and are increasingly being supported by commercial database systems. Materialized views are especially attractive in data warehousing environments because of the query intensive nature of data warehouses. [3] Typical Data warehouse queries are complex and ad-hoc in nature and normally these queries access huge volumes of warehouse data and perform many joins and aggregations. Query response time and throughput are therefore more important than transaction throughput. The data warehousing environment provides a computerised interface that enables business decision-makers to creatively approach, analyse and understand business problems. The aim of the data warehouse system is to turn data into strategic decision making information and to bring solutions to users. This process is done by tuning the data at many steps.[2]

Materialized view eliminates the overhead associated with expensive joins and aggregations for a large or important class of queries. Queries to large databases often involve joins between tables, aggregations such as average, sum, count or both aggregation & joins.

Materialized views can provide massive improvements in query processing time, especially for aggregation queries over large tables.[6] A materialized view takes a different approach in which the query result is cached as a concrete table that may be updated from the original base tables from time to time. This enables much more efficient access, at the cost of some data being potentially out-of-date. It is most useful in data warehousing scenarios, where frequent queries of the actual base tables can be extremely expensive.

In addition, because the materialized view is manifested as a real table, anything that can be done to a real table can be done to it, most importantly building indexes on any column, enabling drastic speedups in query time. In a normal view, it's typically only possible to exploit indexes on columns that come directly from (or have a mapping to) indexed columns in the base tables; often this functionality is not offered at all.

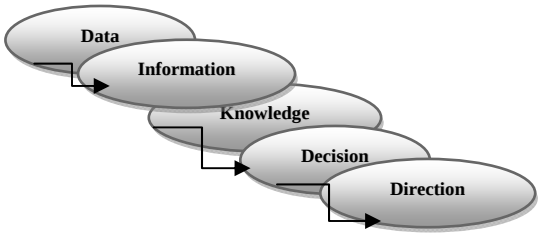
A multidimensional data warehouse (MDW) is a repository in which data is organized along a set of dimensions D = d1, d2…. dn. A possible way to design a MDW is the star-schema in which, for each dimension, there is a dimension table Di that has di as its primary key and also uses a fact table. [3]
Materialized views were implemented first by the Oracle database [9]. These storage structures have attracted much attention since then. The life cycle of MVs have three major stages:

**View design**: determining what views to materialize, including how to store and index them.
**View maintenance**: efficiently updating materialized views when base tables are updated.
**View exploitation**: making efficient use of materialized views to speed up query processing. [8]

**Creation and Maintenance**
Data warehouses contain large amounts of information, often collected from a variety of independent sources. Decision support functions in a warehouse, such as on-line analytical processing (OLAP), involve hundreds of complex aggregate queries over large volumes of data. It is not feasible to compute these queries by scanning the data sets each time, Warehouse applications therefore build a large number of summary tables, or materialized aggregate views, to help them increase the system performance. [5]

It is a matter of high concern to decide what data is to be depicted in materialized views and in what numbers materialized views should be created. This decision is obviously influenced by the pattern users and applications access the data. It may happen, if not properly given attention, that a materialized view is created but the data depicted in that materialized view is never or rarely accessed. On the other hand, queries may not be redirected to use materialized views, instead, base tables are accessed in query responses.

The design, implementation and maintenance of materialized views is not single time activity, this process will be carried out through out the life of the data warehouse and database systems. The DBA periodically checks whether a materialized view is in use or not. If a materialized view is not in use, it is dropped to achieve space and time efficiency. New Materialized Views need to be created incase if more data extraction queries are accessing base tables. Various queries that require data from one domain can easily be answered by creating a relevant MV. There are various algorithms in this regards. Like MiniCon Algorithm, it is a scalable algorithm for answering queries using views. [7]




(Figure1: Data Analysis Chart)

Types of Materialized Views:

The types of materialized views used are as follows:

**a) Materialized views with aggregates:**

Materialized views contain aggregates in data warehouses for fast refresh to be possible. The valid aggregates functions are Sum, Count, Average, Variance, Min, Max, Standard Deviation etc

**b) Materialize Views with Joins:**

Some materialized views contain only joins and no aggregates. The advantage of creating this type of view is that expensive joins will be pre-calculated. A materialized view containing only joins can be defined to be refreshed On Commit or On Demand.

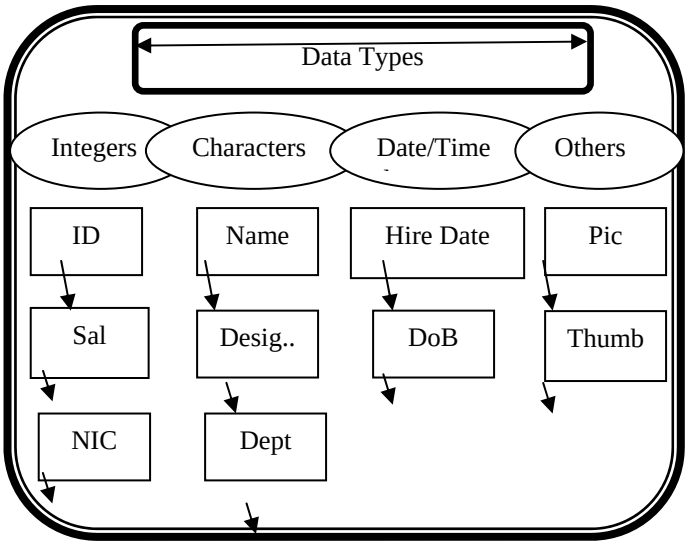
**c) Nested Materialized Views:**
A nested materialized view is that type of materialized view whose definition is based on another materialized view. A nested materialized view can reference other relations in the database in addition to referencing materialized views. [9]

**2. Child-View Creation**

Once a base-relation or a materialized view is created and populated, then the process of creating child MVs is started. In the first step the total number of columns is determined. Afterwards, the data types of all columns are determined. For this purpose data dictionary of the database is retrieved.

Data types of the columns are also important in deciding the kind of aggregation operation on the whole data set. Types of values in a column are first scanned whether these values can be used in aggregation or it can be made part of a subset in the form a child materialized view. It is a known fact that numeric data can easily be aggregated.

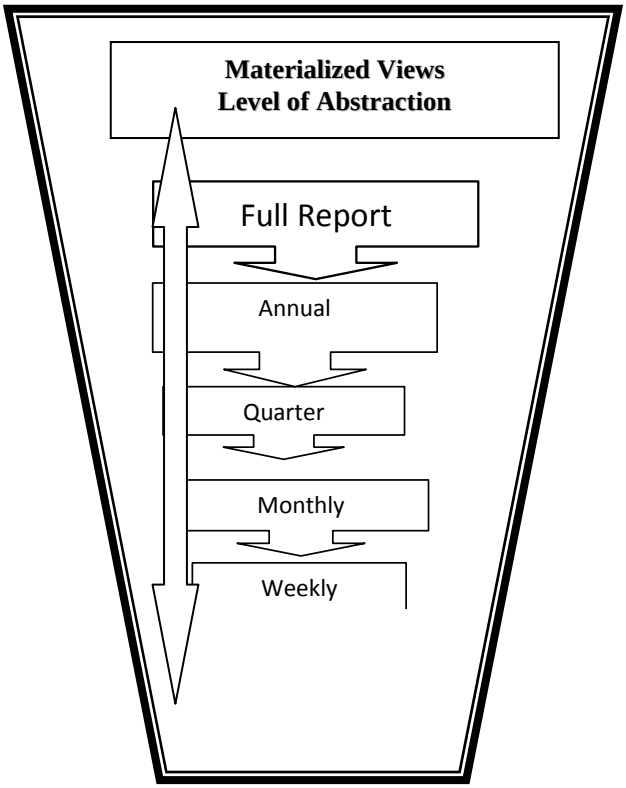
But there are certain values which can not be aggregated at all, for example, columns containing IDs can not be used in such context. We focus text databases in this work, because objects of various kinds that can be referenced from inside a database can not be aggregated. Images, audio or video files which are referenced from inside the database can not be aggregated or combined together as we aggregate numeric data using AVG or SUM operation. Character data again depends whether it can be used in aggregation or not. Like if *names* are stored in a column, so this column can not be summarized in any manner considering the values it contain. *Name* column supports no aggregation function excepts COUNT. But if we take *address*, it can be taken for aggregation based on House, Street, Sector, Town, City, Country and Region. *address* supports partial aggregation as it can be used to aggregate the data but standard functions of aggregation like SUM, AVG can not be performed on it. And if we consider, that means some string/character columns inherently have hierarchies while others do not. In figure 2 we have shown different data type's columns, by the combination of these we can create our child Materialized views.




(Figure 2: Aggregation Levels)

We need to look for columns where hierarchies can be used as a tool to aggregate data. This can be done by observing the dataset manually as the data dictionary can not be used to have information regarding each column about its aggregation. In a typical DBMS data dictionary, no such information is found nor there do any space where the schema designer can add this aggregation information. Furthermore, in a dimensional context, the data warehouse schema is variable which may lead to a bit of performance overhead in maintaining such information. However, a subschema specific to a base MV is created in this work to

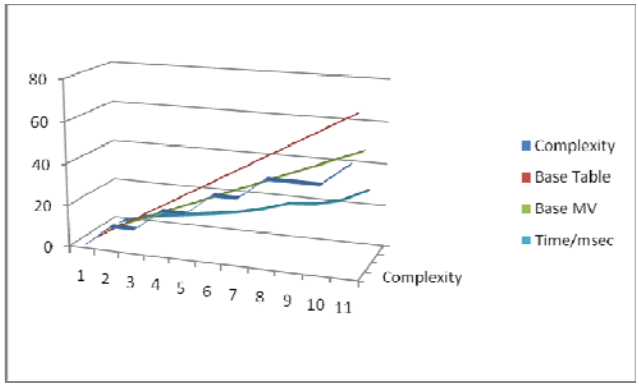
have information regarding each summarizable / aggregable column. Then the kind of aggregation function is decided for each column. If the column is numeric then standard group function may be used for aggregation and if it is a string (e.g. *address)* then contents of the value may be used for aggregation and if it is a DATE/TIME column then only compatible aggregation will be performed. Then the level of aggregation is also decided i.e. from the most granular to the most general level. This information will be used in context with other partner columns in a potential child MV. All such information result in an AGGREGATION PLAN of the MV.

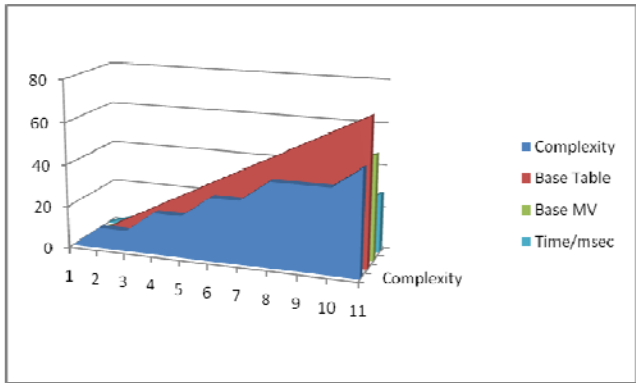



(Figure 3: MV Level of Abstraction)

Once an *aggregation plan* is finalized, then it can be implemented. We have shown Level of aggregations in figure 3. A table containing $d$ columns can have $2^d$ various combinations of columns. In aggregation, the number of combinations of columns may be double of this because here content based aggregation is also performed. This will result in too many child MVs which will be surely unmanageable and unusable. For this either the aggregation plan is again observed and those potentially useful candidate MVs may be kept and rest of the MVs are deleted. In order to answer a query, a data integration system needs to translate a query formulated on the mediated schema into one that refers directly to the schemas in the data sources. [5]

The following chart shows a comparison of time vs queries complexity of base relation to MV tree.




| S# | Complexity | Base Table | Base MV | Time/m sec |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 2 | 10 | 7 | 5 | 3.78 |
| 3 | 10 | 14 | 10 | 5.87 |
| 4 | 20 | 21 | 15 | 7.65 |
| 5 | 20 | 28 | 20 | 9.39 |
| 6 | 30 | 35 | 25 | 11.36 |
| 7 | 30 | 42 | 30 | 14.38 |
| 8 | 40 | 49 | 35 | 18.36 |
| 9 | 40 | 56 | 40 | 19.32 |
| 10 | 40 | 63 | 45 | 22.39 |
| 11 | 50 | 70 | 50 | 28.39 |

(Chart 1:
Complex Queries Vs Time Series for Base Relation & MV Tree)

## 3. Future Work

The automatic creation of child materialized view is really a complex task. There are a lot of things needed to be considered. While working on this paper, we have found out areas where we think improvements can be brought in. The data dictionary of a database or data warehouse needs modification to include meta data for deciding which numeric and string attributes can be exploited for creating new MVs. Attributes containing multimedia data are not in the scope of this paper. However, our work can be extended by including multimedia data operations as well in MV child creation plan. The decision regarding the number of child materialized view also difficult to answer, so one can also find a balanced number of MVs with high access frequency and remove the rest of the MVs

**Conclusion**

Materialized Views do contribute in answering queries efficiently to improve the over all performance of Data Warehouse. Efficient query answering can further be speeded up by creating various child materialized views. Data extraction queries select the best MV which can fulfill its data requirements. For creating child MVs, initially the data types of fields/columns of the base MV are determined. All of the columns are grouped according to the data types. Incase of numeric data types, those columns are separated which can not be aggregated e.g, ID, SSN, from numeric attributes where various aggregation operations can be carried out. This whole process is carried out for string attributes as well. This whole process result into an aggregation plan, which later on is translated into a script and is then executed. The process of query answering is made efficient by having more MVs having the potential data required to fulfill maximum requirements of query.

**References**

[1] *"New Trends in Data Warehousing and Data Analysis"*Series: Annals of Information Systems, Vol. 3 Kozielski, Stanislaw; Wrembel, Robert (Eds.) 1st Edition. 2nd Printing. 2009, XVIII, 364 p. 152 illus.

[2] *"Advances and Research Directions in Data Warehousing Technology"*
Australasian Journal of Information Systems, Vol 7, No 1 (1999)
http://dl.acs.org.au/index.php/ajis/article/view/287

[3] *"A Case for Dynamic View Management"*
Yannis Kotidis AT&T Labs Research and Nick Roussopoulos University of Maryland
ACM Transactions on Database Systems, Vol. 26, No. 4, December 2001, Pages 388–423.

[4] *"Answering queries using views: A survey"*
A.Y. Halevy Department of Computer Science and Engineering, University of Washington, Seattle,WA, 98195

[5] *"Data Cubes and Summary Tables in a Warehouse*" Inderpal Singh Mumick, Dallan Quass, Barinderpal Singh Mumick, *"Maintenance of* June 1997 ACM SIGMOID RECORD.

[6] *"Materialized View Selection and Maintenance Using Multi Query Optimization"*
Hoshi Mistry, Prason Roy, S.Sudershan, K.Ramamritham
ACM SIGMOD 2001 May 21-24, Santa Barbara, California USA

[7] *"MiniCon: A scalable algorithm for answering queries using views"* Rachel Pottinger, Alon Halevy University ofWashington, Department of Computer Science and Engineering, Box 352350 Seattle,WA 98195, USA

[8] *"Optimizing Queries Using Materialized Views: A Practical, Scalable Solution"*
Jonathan Goldstein and Per-Åke Larson
Microsoft Research, One Microsoft Way, Redmond, WA 98052

[9] "Oracle 9i" Data Warehousing Guide, Release 2 (9.2), March 2002, A-96520-01

[10] *"Oracle Database 11g for Data Warehousing and Business Intelligence" Oracle Publishers.*

# Ontology Based Feature Driven Development Life Cycle

**Farheen Siddiqui[1], M. Afshar Alam[1]**

**[1] Department of Computer Science, Hamdard University, New Delhi -110025 India**

**Abstract**

The upcoming technology support for semantic web promises fresh directions for Software Engineering community. Also semantic web has its roots in knowledge engineering that provoke software engineers to look for application of ontology applications throughout the Software Engineering lifecycle. The internal components of a semantic web are "light weight", and may be of less quality standards than the externally visible modules. In fact the internal components are generated from external (ontological) component. That's the reason agile development approaches such as feature driven development are suitable for application's internal component development. As yet there is no particular procedure that describes the role of ontology in FDD processes. Therefore we propose an ontology based feature driven development for semantic web application that can be used form application model development to feature design and implementation. Features are precisely defined in the OWL-based domain model. Transition from OWL based domain model to feature list is directly defined in transformation rules. On the other hand the ontology based overall model can be easily validated through automated tools. Advantages of ontology-based feature Driven development are also discussed.

***Keywords: S****emantic Web , Feature Driven Development ,Agile Development.*

## 1. Introduction

The Software Engineering and Knowledge Engineering groups work on overlapping domain. Software Engineering people pay more attention to software modeling and Knowledge Engineering community has come up with variety of modeling approaches in order to realize the vision of the semantic web [1].Semantic web has made this overlap even more wide but still there is less forums for discussing synergies is (e.g. SWESE1, SEKE2 and W3C3) .The methods on integrating Software and Knowledge Engineering approaches focus on approaches of meta-modeling, but are abstract for software engineers in terms of there application in software process. Current approaches of modeling only partially solve the problem related to component reuse, composition, validation, information and application integration, software testing and quality. Such basic needs are generating new approaches towards every single aspect in software engineering.

Domain analysis is an essential activity for successful reuse across applications in the same domain. Domain model is essential for domain and application-specific development. And therefore should meet some requirements. First, it should provide guidance for the design of architecture and components. Second, the model should provide means to get validated against system constraints. Third, it should be customizable for specific application. In "semantic web" era, developer would discover shareable domain models and knowledge bases from a variety of interrelated repositories and then connect them together with application specific components. Thus all applications that share overlapping domain models would then have a certain degree of interoperability built in. These sharable domain models are referred as domain ontology and provide many benefits such as model reuse, flexibility, consistency checking and reasoning. Also new technologies and tools have been developed for ontology representation, machine-processing, and ontology sharing. This makes their adoption in real-world applications much easier. While ontologies are about to enter mainstream Software Engineering practices, their applications in software engineering are manifold. Despite of using a well defined domain model it is not uncommon for software projects to exceed budget, blow schedule, and deliver something less than desired .The main reason behind this is the scenario of ever changing user requirement and lack of communication between customer and developer team. Therefore a process for delivering frequent, tangible, working results is most desired. Agile development approaches focus on these issues and feature driven development is among one of the approaches towards it. The remainder of this paper is organized as follows. Section 2 presents the ontology based feature driven development and the stages involved in it. Section 3 introduces the ontology-based "overall model", Section 4 elaborates the process of feature list development and planning and section 5 discuss about component

development. Finally, we draw our conclusions with discussion of ontology-based feature modeling and future work in section 6.

## 2. Feature Driven Development

Feature Driven Development is a model-driven short-iteration process. It begins with establishing an "overall model" shape. Then it continues with a series of two week "design by feature, build by feature" iterations. The features are small "useful in the eyes of the client" results. Iteration like "build the admission subsystem" would take too long to complete. Iteration like "build the data access layer" is not exactly client-valued. In contrast, a small feature like "assign unique enrollment number" is both short and client-valued. FDD is based on its first process of developing overall model. This process is so critical that it is referred as process 1 in FDD life cycle. Therefore a strictly defined modeling basis for "overall model" is essential, which should provide a mechanism to connect model elements in various development phases. Ontology related theory is a suitable way to achieve our goals. Ontology is a conceptualization of a domain or subject area typically captured in an abstract model of how people think about things in the domain [2]. Rubén [2] considers domain models as narrow or specialized ontology, and the main difference is that domain models define abstract concepts in an informal way and have no axioms. Because of the facilities for the generalization and specialization of concepts and the unambiguous terminology it provides [3], ontology has been widely used in domain knowledge representation and requirement modeling, reuse and consistency checking. For example, Sugumaran etc. [4] proposed a semantic-based approach to component retrieval, in which ontology and domain models are adopted for capturing application domain specific knowledge to express more pertinent queries for component retrieval. Girardi etc. [3] proposed GRAMO, an ontology based technique for the specification of domain and user models in Multi-Agent domain engineering.

The purpose of this paper is to reduce the gap between knowledge engineering and software engineering by using ontology in every step of a FDD process.This paper proposes an ontology-based feature driven development methodology, in which OWL ontology is considered as an overall model and is used at every step of FDD. In this way, we can provide better support for domain modeling, and succeeding domain design and implementation. First, ontology-based feature model can be formally represented easily and validation of the model can be realized through ontology reasoning. Second, the ontology-based unambiguous terminology provide precise and detailed semantic knowledge for the domain, so the ontology based feature model can also be adopted as the domain business model and contain enough information for component description and architecture design.

## 3. Ontology Based Feature Driven Development

In this section, we will present method of using ontologies in the context of FDD. The presentation will be in the order of FDD life cycle as described in fig 1. In each step we will discuss how ontology can be used and what benefits we can achieve by its usage.

Traditionally FDD life cycle is based on following five processes.:

Process #1: Develop an overall model (using initial requirements/ features, snap together with components, focusing on shape).

Process #2: Build a detailed, prioritized features list.

Process #3: Plan by feature.

Process #4: Design by feature (using components, focusing on sequences).

Process #5: Build by feature.

For ontology based feature driven development we have merged these into three stages as depicted in Fig 1. At each stage ontology is used as the basic building block. Software modeling languages and methodologies can benefit from the integration with ontology languages such as RDF and OWL in various ways, e.g. by reducing language ambiguity, enabling validation and automated consistency checking Ontology languages provide better support for logical inference, integration and interoperability than MOF-based languages. UML-based tools can be extended more easily to support the creation of domain vocabularies and ontologies. Since ontologies promote the notion of identity, Ontology Definition Metamodel and related approaches simplify the sharing and mediation of domain models. Since a domain model is initially unknown and changes over time, a single abstraction and separation of concerns is considered feasible if not necessary Therefore a single representation of the domain model should be shared by all participants throughout the lifecycle to increase quality and reduce costs. The mapping of a domain model to code should therefore be automatized to enable the dynamic use by other components and applications. Fig 1 depicts the three dimensional view of FDD life cycle that uses semantic web technologies at each stage of development. FDD begins with application model development. We use OWL and SWRL to define the entities, classes, hierarchies and domain rules in form of problem ontology. At second process the feature list is generated from the problem ontology and planning is done with SQWRL[15].At final process of component building each feature generated

from problem ontology is designed and implemented using APIs like jena.In the following

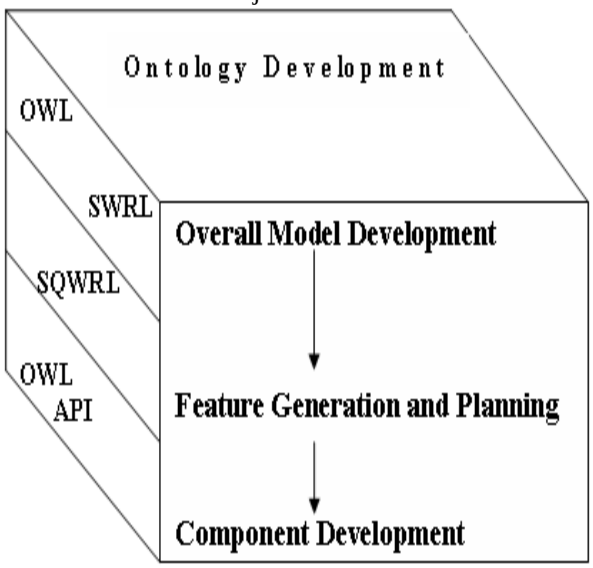



Fig. 1 FDD Life Cycle Model.

## 4. Develop an Overall Model

Developing the initial model shape needs involvement of both domain and development members. Domain member starts with presenting an abstract view and scope of the system within application context. The domain and development members develop a rough model that can be followed at the initial stage. Later on the domain and development member stepwise explores each detail aspect of the system and merge the understanding in the initial model alongside adjusting model shape. The development of overall model starts when the client is ready to proceed with the building of a system but he might not express the requirement in any concrete format. Hence at first this phase deals with gathering the desired system functionality from the customers. Since the involved software engineers are often no domain experts, they must learn about the problem domain from the customers. A different understanding of the concepts involved may lead to an ambiguous, incomplete specification and major rework after system implementation. Therefore it is important to assure that all participants in the phase have a shared understanding of the problem domain. Moreover, change of requirements needs to be considered because of changing customer's objectives.

An ontology can be used for both, to describe requirements specification documents [5, 6] and formally represent requirements knowledge [7,8]. Ontologies can cover semi-formal and structured as well as formal representation [7]. Further, the "domain model" represents the understanding of the domain under consideration, i.e. in the form of concepts, their relations and business rules. It is formalized using a conceptual modeling language such as the UML. Moreover, the problem domain can be described using an ontology language, with varying degrees formalization and expressiveness. In contrast to traditional knowledge-based approaches, e.g. formal specification languages, ontologies seem to be well suited for an evolutionary approach to the specification of requirements and domain knowledge [7] that is needed to achieve agility in development cycle. Moreover, ontologies can be used to support requirements management and traceability [6]. Automated validation and consistency checking are considered as a potential benefit compared to semi-formal or informal approaches providing no logical formalism or model theory. Finally, formal specification may be a prerequisite to realize model-driven approaches in the design and implementation phase. At the end of the process 1, an overall ontology based model is developed and based on that an informal feature list is noted down. In this paper to support the life cycle, we have taken an example of University system. Following the above procedure the developer and domain expert build Education ontology in Protégé .Fig 2 shows graphical representation of Education ontology developed in Protégé.

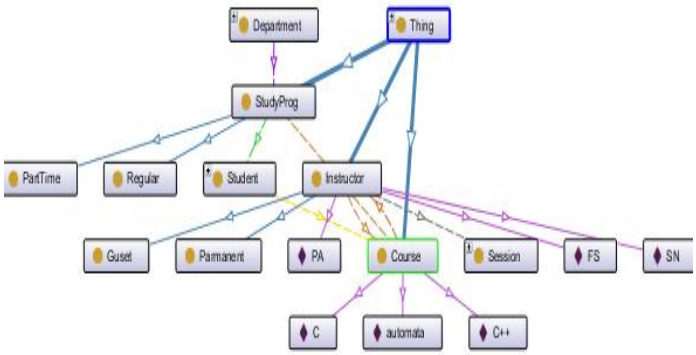



Fig. 1 Education Ontology in Protégé.

Ontologies are purely conceptual models that capture domain concepts and neglects domain-restricted rules. If the requirements model violate these rules or contradict the usual business behavior, they become unreasonable. We have used SWRL to model the integrity rules and derivation rules which restrict the business behavior. The

requirements model represented by domain ontology can be checked for consistency using HermiT reasoners and rules can be checked with Jess rule Engine.Thus model development process is both guided by domain ontology and restricted by domain rules. Therefore, the model would comply with both business needs and domain knowledge. A rule contains one or more antecedent and one consequent, the description is as follows:

```
<owlr : rule rdf : ID =″rule ID″>
<owlr : annotation >ruleName</ owlr : annotation >
<owlr :body>antecedent</ owlr :body>
<owlr :head>consequent</ owlr :head>
</ owlr : rule>
```

Consider the rule that a person having a salary associated is an employee. This rule can be ststed in SWRL as:

```
Person(?p) ^ salary(?p, ?s) -> Employee(?p)
```

and encoded in ontology as

```
DLSafeRule>
<Body>
<ClassAtom>
<Class IRI="#Person"/>
<Variable IRI="urn:swrl#p"/>
</ClassAtom>
<DataPropertyAtom>
<DataProperty IRI="#salary"/>
<Variable IRI="urn:swrl#p"/>
<Variable IRI="urn:swrl#s"/>
</DataPropertyAtom>
</Body>
<Head>
<ClassAtom>
<Class IRI="#Employee"/>
<Variable IRI="urn:swrl#p"/>
</ClassAtom>
</Head>
    </DLSafeRule>
```

## 6. Feature Generation and Planning

While building the feature list, the main task is to identify the features, groups them hierarchically, prioritizes them, and weights them. In subsequent iterations of this process, smaller teams tackle specialized feature areas. We propose to establish one to one correspondence between the ontology and the feature list development. We can use the ontology develop at previous step to generate features supported by it and can also group features into feature set.

### 5.1 Feature List Generation

The process starts with the informal features list from FDD Process 1. It then:

_ transforms object property in the ontology into features of their domain,

_ transforms classes in the ontology into feature sets

We can use the formats:

_ For features: <action> the <object property-range> <by|for|of|to> a(n) <Class-name>

_ For feature sets: <Class-name> module including all subclass of <Class-name>

_ For major feature sets: <ontology-name> management

For example in Education Ontology classes can be Student , Department ,StudyProgram, Courses, Session ,Attendance ,Instructor etc.Also a object property hasStudyProg has domain of Department and range of StudyProg. Therefore a feature: department offers study program can be considered as a feature in form

Offering of StudyProgram by Department, which is a triple of form:

<action> <object property range> by|for|of|to <object property domain>

This can be inferred from ontology as hasStduyProg is a object property of Department and this feature belongs to department module of education management. To exit this process, the features-list team must deliver a detailed features list, grouped into major feature sets and feature sets.

### 5.2 Feature Planning

At planning stage the project manager, the development manager, and the chief programmers establish milestones The planning team determines the development sequence and sets initial completion dates for each feature set and major feature sets for “design by feature, build by feature” iterations. Using the development sequence and the feature weights as a guide, the planning team assigns classes to class owners. Using the development sequence and the feature weights as a guide, the planning team assigns chief programmers as owners of feature sets (classes in ontology). Every class in ontology can be associated with a property of “hasowner”.A feature indicates the class(es) involved and a query can be framed in SQWRL to fetch the class owner of corresponding classes in ontology. For example to find out owner of a particular class Instructor for feature “assign Course to Instructor” the following query can be used:

Course(?c) ^ Instructor(?I) ^ hasCourse(?I, ?c) ^ hasOwner(I,P)-> sqwrl:select(?I, ?P)

To exit this process, the planning team must produce a development plan, subject to review and approval by the

development manager and the chief architect. The plan consist of an overall completion date, for each major feature set, and feature: its owner and its completion date , for each class, its owner.

## 4. Component Development

This stage consists of iterations feature design, feature implementation.

### 5.2 Feature Design

A chief programmer takes the next feature, identifies the classes likely to be involved, and contacts the corresponding class owners. This feature team works out a detailed sequence diagram. Chief programmer identifies the classes likely to be involved in the design of this feature and identifies the developers needed to form the feature team. He contacts those class owners, initiating the design of this feature.While developing the design the team also can look for components that already exist when implementing functionality, since reuse can avoid rework, save money and improve the overall system quality. Usually, this search for reusable components takes place after the analysis phase, when the functional requirements are settled [9]. Ontologies can help here to describe the functionality of components using a knowledge representation formalism that allows more convenient and powerful querying [10]. One approach implementing this is the KOntoR system that allows storing semantic descriptions of components in a knowledge base and running semantic queries on it. Compared to traditional approaches, ontologies provide two advantages in this scenario. First, they help to join information that normally resides isolated in several separate component descriptions. Second, it provides background knowledge that allows non-experts to query from their point of view .

### 5.2 Feature Implementation

Each class owner builds his object property for the feature. He extends his class-based test cases and performs class-level (unit) testing. Once the code is successfully implemented and inspected, the class owner checks in his class(es) to the configuration management system. When all classes for this feature are checked in, the chief programmer promotes the code to the build process.
At the end of this phase, the feature team must delivers implemented and inspected classes and properties with unit testing. The mapping of a domain model to code should be automated to enable the dynamic use by other components and applications. The programmatic access of ontologies and manipulation of knowledge bases using ontology APIs requires special knowledge by the developers. Therefore an intuitive approach for object-oriented developers is desirable [cf. 23]. This can be achieved by ontology tools that generate an API from the ontology, e.g. by mapping concepts of the ontology to classes in an object oriented language. The generated domain object model can then be used managing models, inferencing, and querying. Tools supporting those features are already available today, e.g. [12] and [13].The domain model encoded in OWL can be used at implementation time with OWL API.
Semantic Web applications usually need to make some ontological commitments, i.e., they need to have hard-coded knowledge about a certain domain ontology. In the example above, the application has hard-coded behavior that depends on the education.owl ontology, which contains classes like **Instructor** and **Course**. The application can exploit reasoning engines like Racer or rule engines like SWRL to expose "intelligent" behavior. All of this is controlled by some logic (in this example it is Java code), which also interacts with the end user by means of interface technologies like JSPs, Swing applications, or Web Services. Protege-OWL API features can be used for developing stand-alone applications. Such applications can load ontologies from the Semantic Web, perform queries on them, add or edit resources from the ontology, classify instances and classes, and write out resulting ontologies to a file. From an object-oriented perspective, Owl API can generate code for class such as:

```
public interface Person {
    String getFirstName();

    void setFirstName(String value);
    ...

}
so that we can use code like this:
public interface Person extends
OWLIndividual {

    String getFirstName();
    void setFirstName(String value);
    ...
}
```

and then provide a default implementation like the following scheme:

```
public class DefaultPerson extends
DefaultOWLIndividual implements Person
{

    public DefaultPerson(KnowledgeBase
kb, FrameID id) {
        super(kb, id);
    }

    public String getFirstName() {
```

```
        RDFProperty property =
getOWLModel().getRDFProperty("firstName
");
        return (String)
getPropertyValue(property);
    }

    public void setFirstName(String
value) {
        RDFProperty property =
getOWLModel().getRDFProperty("firstName
");
        setPropertyValue(property,
value);
    }

    ...
}
```

## 7. Conclusions

A strictly-defined formal basis is essential for applicable domain modeling. In this paper, ontology is used as the foundation of the FDD life cycle. Ontology has been widely adopted in domain knowledge modeling and has corresponding modeling language, such as OWL.Furthermore, rule-based reasoning can be performed on the ontology model for model validating. Establishing a mapping between domain model and the architecture is the objective of domain engineering [14]. However, there is a large gap between the domain model representation and actual implementation. We can reduce the gap by establishing a smooth transition from elements in the domain model (i.e. features) to elements in the architecture model (i.e. components). In our approach, domain ontology (i.e. the ontology-based overall model) is also representation basis for component semantics. Our future work will be based on the complete implementation of an education system through feature driven development using education ontology. Also in future we will develop an ontology based architecture and design pattern for semantic web application.

## References

1. Berners-Lee, T., Hendler J. and Lassila, O.: The Semantic Web. Scientific American 284(5) (2001)

2. Rubén Prieto-Díaz. A faceted approach to building ontologies. Proceedings of IEEE International Conference on Information Reuse and Integration (IRI 2003). 2003: 458~465.

3.Rosario Girardi, Carla Gomes de Faria. An ontology-based technique for the specification of domain and user models in multi-agent domain. CLEI electronic journal, Vol.7(1), 2004.

4. Vijayan Sugumaran, Veda C. Storey. A semantic-based approach to component retrieval. ACM SIGMIS Database, Vol.34:pages 8-24, 2003.

5. Mayank, V., Kositsyna, N., Austin, M.: Requirements Engineering and the Semantic Web, Part II. Representation, Management, and Validation of Requirements and System-Level Architectures. Technical Report. TR 2004-14, University of Maryland (2004)

6. Decker, B., Rech, J., Ras, E., Klein, B., Hoecht, C.: Selforganized Reuse of Software
Engineering Knowledge supported by Semantic Wikis. In: Proc. of Workshop on Semantic Web Enabled Software Engineering (SWESE). November (2005)

7 Lin, J., Fox, M. S.; Bilgic, T.: A Requirement Ontology for Engineering Design. Enterprise Integration Laboratory,, University of Toronto, Manuscript, September (1996)

8. Wouters, B., Deridder, D., Van Paesschen, E.: The Use of Ontologies as a Backbone for Use Case Management. In: "European Conference on Object-Oriented Programming (ECOOP 2000), Workshop : Objects and Classifications, a natural convergence" (2000)

9. Cheesman, J. and Daniels, J.: UML Components: A Simple Process for Specifying Component- Based Software. Addison-Wesley, 2000.

13. Mili, A., Milli, R.., Mittermeir, R.T.: A Survey of Software Reuse Libraries. In: Annals of Software Engineering, vol. 5, (1998) 349-414

10. Happel, H.-J., Korthaus, A., Seedorf, S., Tomczyk, P.: KOntoR: An Ontology-enabled Approach to Software Reuse. In: Proc. of the 18th Int. Conf. on Software Engineering and Knowledge Engineering (SEKE), San Francisco, July (2006)

11. Knublauch, K.: Ramblings on Agile Methodologies and Ontology-Driven Software
Development. In: Proc. of the Workshop SWESE, ISWC, Galway, Ireland (2005)

12. Knublauch, H., Oberle, D., Tetlow, P., Wallace, E.: A Semantic Web Primer for Object-Oriented Software Developers. W3C Working Group Note, http://www.w3.org/TR/sw-oosdprimer/, 9 March (2006)

13. Völkel, M.: RDFReactor - From Ontologies to Programatic Data Access. In: Proc. of the Jena User Conference 2006. HP Bristol, May (2006)

14 Kyo C Kang , Sajoong Kim , Jaejoon Lee , et al. FORM: A Feature-Oriented Reuse
Method with Domain-Specific Reference Architectures. Annals of Software Engineering,
1998,5 :143～168.

15. O'Connor, M.J. and Das, A. "SQWRL: a Query Language for OWL" OWL: Experiences and Directions (OWLED 2009), Fifth International Workshop, Chantilly, VA, 2009.

# Hole Detection for Increasing Coverage in Wireless Sensor Network Using Triangular Structure

**Shahram Babaie[1] and Seyed Sajad Pirahesh[2]**

**[1] Department of Computer engineering, Tabriz-Branch, Islamic Azad University, Tabriz, Iran**

**[2] Department of Computer engineering, Tabriz-Branch, Islamic Azad University, Tabriz, Iran**

**Abstract**

The emerging technology of wireless sensor network (WSN) is expected to provide a broad range of applications, such as battlefield surveillance, environmental monitoring, smart spaces and so on. The coverage problem is a fundamental issue in WSN, which mainly concerns with a fundamental question: How well a sensor field is observed by the deployed sensors? Mobility is exploited to improve area coverage in a kind of hybrid sensor networks. The main objective for using mobile sensor nodes is to heal coverage holes after the initial network deployment, when designing a hole healing algorithm, the following issues need to be addressed. First, how to decide the existence of a coverage hole and how to estimate the size of a hole. Second, what are the best target locations to relocate mobile nodes to repair coverage holes? We use the triangular oriented diagram (HSTT) for aim to goal where its simple ,have low calculation among construction and it is great to calculate the size of hole exactly .

***Keywords:*** *Wireless Sensor network; Area Coverage; hole detection; size calculation; target location; Coverage.*

## 1. Introduction

Recent advances in micro-electro-mechanical systems, embedded processors, and wireless communications have led to the emergence of Wireless sensor networks (WSNs), which consist of a large number of sensing devices each capable of sensing, processing and transmitting environmental information. Applications of WSNs include battlefield surveillance, environmental monitoring, biological detection, smart spaces, industrial diagnostics, and so on [1].

A fundamental issue in WSNs is the coverage problem [2, 3]. The coverage problem is heavily dependent on the coverage model of individual sensor and the locations of the deployed sensor nodes. Sensor coverage model can be considered as a measure of the quality of service of sensor's sensing function and is subject to a wide range of interpretations due to a large variety of sensors and applications. In the literature, a widely used sensor coverage model is the sensing disk model where a sensor can cover a disk centered at itself with a radius equal to a fixed sensing range. Network sensing coverage on the other hand can be considered as a collective measure of the quality of service provided by sensor nodes at different geographical locations. In many cases, we may interpret the coverage concept as a non-negative mapping between the space points (of a sensor field) and the sensor nodes (of a deployed sensor network). For example, given the sensing disk model, the area (space points) covered by a set of sensors is the union of their sensing disks.

Wireless sensors can be either deterministic placed or randomly deployed in a sensor field. Deterministic sensor placement can be applied to a small to medium sensor network in a friend environment. When the network size is large or the sensor field is remote and hostile, random sensor deployment might be the only choice, e.g., scattered from an aircraft. It has been shown that a critical sensor density exists beyond which a sensor field can be completely covered almost surely in every random deployment [4, 5]. To guarantee complete coverage in one random deployment, it is often assumed that the number of scattered sensors is more than that required by the critical sensor density. However, this normally requires a great number of sensor nodes to be deployed another way to improve network coverage is to leverage mobile sensor nodes. Mobile sensor nodes are equipped with locomotive platforms and can move around after initial deployment, for example, the mobile sensor nodes Robomote [6] and iMouse [7]. Although in general a mobile sensor node is more expensive than its stationary compeer, it can serve much functionality such as a data relay or collector, and can greatly improve many network performances such as enhancing timeliness of data report. In this article, our focus is to healing coverage hole using genetic algorithm with minimize total movement of mobile sensor.

## 2. RELATED WORK

Bang Wang, Hock Beng Lim, Di Ma with article [12] proposed:

### 2.1 Hole detection and hole size estimation

Voronoi diagram can be used to detect a coverage hole and calculate the size of a coverage hole [8, 9]. A Voronoi diagram for N sensors $s_1$, $s_2$,…,$s_N$ in a plane is defined as the subdivision of the plane into N cells each for one sensor, such that the distance between any point in a cell and the sensor of the cell is closer than that distance between this point and any other sensors. Two Voronoi cells meet along a Voronoi edge and a sensor is a Voronoi neighbor of another sensor if they share a Voronoi edge. We refer the reader to [10] for more discussions on Voronoi diagram and its applications.

A Voronoi diagram is first constructed for all stationary sensor nodes, assuming that each node knows its own and its neighbors' coordinates. Wang et al. [9] proposes a localized construction algorithm to construct a local Voronoi diagram: Each node constructs its own Voronoi cell by only considering its 1-hop neighbors. After the local Voronoi diagram construction, the sensor field is divided into sub regions of Voronoi cells and each stationary node is within a Voronoi cell. A node is a Voronoi neighbor of another one if they share a Voronoi edge. Fig. 2 illustrates a Voronoi diagram in a bounded sensor field, where the boundaries of the sensor field also contribute to a Voronoi cell. According to the property of a Voronoi diagram, all the points within a Voronoi cell are closest to only one node that lies within this cell. Therefore, if some points of a Voronoi cell are not covered by its generating node, these points will not be covered by any other sensor and contribute to coverage holes. If a sensor covers all of its Voronoi cell's vertices, then there are no uncovered points within its Voronoi cell; otherwise, uncovered points exist within its Voronoi cell.

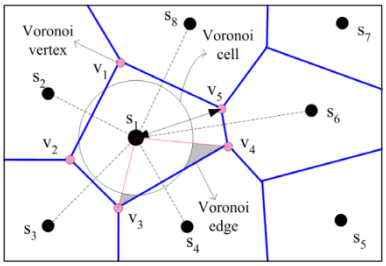



Fig. 1 Illustration of using Voronoi diagram to detect a coverage hole and decide the hole size.

### 2.2 Destination selection

After deciding the existence of a coverage hole and its size, a stationary node needs to decide the number of mobile nodes and the target locations of these mobile nodes to heal its holes. Ghosh [8] proposes that for each Voronoi vertex, one mobile node should be used to heal the coverage hole around this Voronoi vertex, if the size of its coverage hole within the Voronoi cell is larger than a threshold.

Wang et al. [13] convert the sensor movement problem into a maximum weight maximum-matching problem to decide which mobile node should move to which target location.

## 3. OUR WORK

We assume there is an environment like fig 1. That fill with stationary sensors at the first fore implement. In our diagram (HSTT) we connect the center of sensors sensing to gather with condition of made a triangular every three adjacent sensors.

Our diagram can be used to detect a coverage hole and calculate the size of a coverage hole

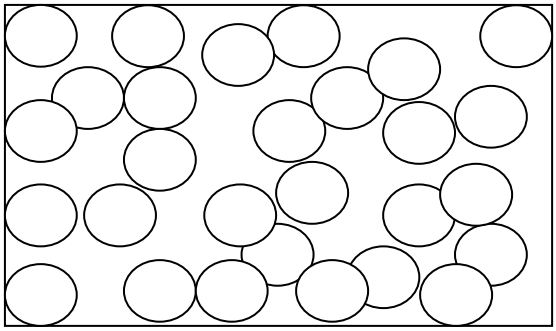


Fig.2 Initial deployment.

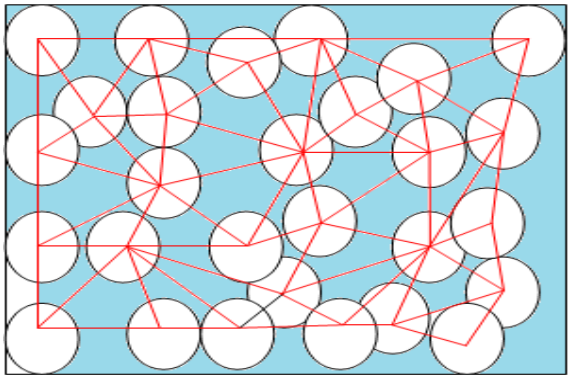


Fig.3 Constraction the triangular oreinted structure.

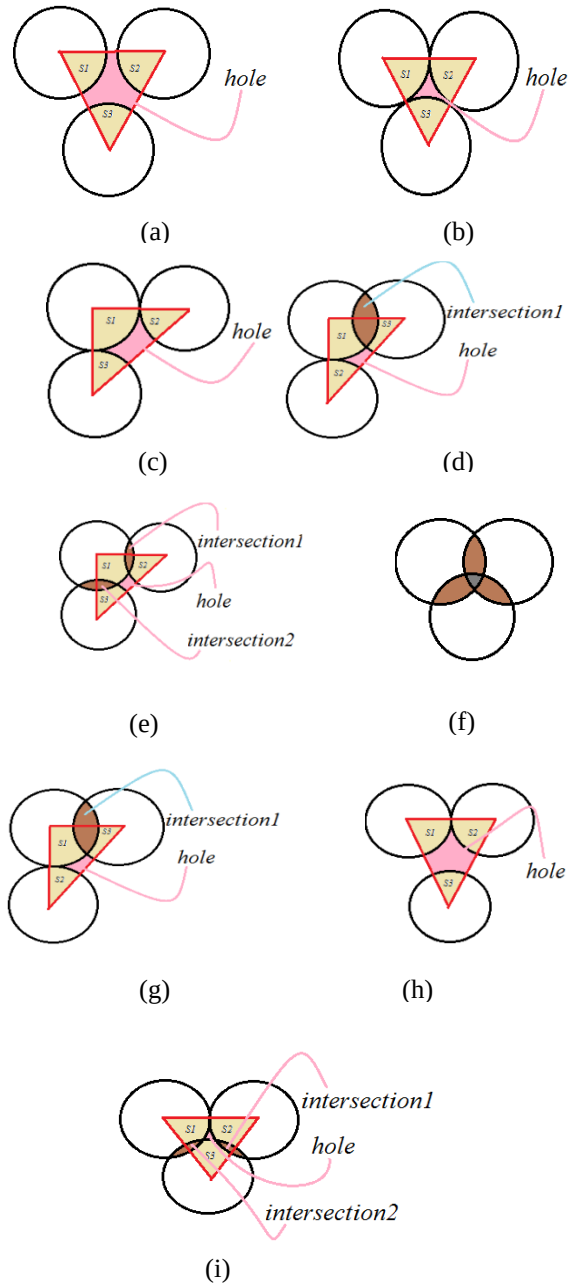



Fig.4 The total possible sates with three sensors.

Relation to area of the triangle.

Let A be the triangle's area and let a, b and c, be the lengths of its sides. By Heron's formula, the area of the triangle is

$$area = A$$

$$= \frac{1}{4}\sqrt{(a+b+c)(a-b+c)(b-c+a)(c-a+b)}$$

$$= \sqrt{s(s-a)(s-b)(s-c)} \quad (1)$$

Where $s = \frac{a+b+c}{2}$ is the semi perimeter.

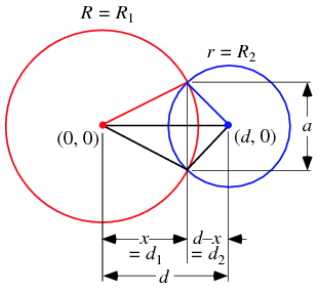



Fig.5 Intersection between two circles

Let two circles of radii **R** and **r** and centered at (0,0) and (d,0) intersect in a region shaped like an asymmetric lens. The equations of the two circles are

$$x^2 + y^2 = R^2 \quad (2)$$

$$(x-d^2) + y^2 = r^2 \quad (3)$$

Combining (1) and (2) gives

$$(x-d)^2 + (R^2 - x^2) = r^2 \quad (4)$$

Multiplying through and rearranging gives

$$x^2 - 2dx + d^2 - x^2 = r^2 - R^2 \quad (5)$$

Solving for **x** results in

$$x = \frac{d^2 - r^2 + R^2}{2d} \quad (6)$$

The chord connecting the cusps of the lens therefore has half-length **y** given by plugging **x** back in to obtain

$$y^2 = R^2 - x^2 = R^2 - \left(\frac{d^2 - r^2 + R^2}{2d}\right)^2$$

$$= \frac{4d^2R^2 - (d^2 - r^2 + R^2)^2}{2d} \quad (7)$$

Solving for **y** and plugging back in to give the entire chord length a=2 then gives

$$a = \frac{1}{d}\sqrt{4d^2R^2 - (d^2 - r^2 + R^2)^2} \quad (8)$$

$$= \frac{1}{2}\sqrt{(-d + r - R)(-d - r + R)(-d + r + R)(d + r + R)}$$

This same formulation applies directly to the sphere-sphere intersection problem To find the area of the asymmetric "lens" in which the circles intersect, simply use the formula for the circular segment of radius $R'$ and triangular height $d'$

$$A(R', d') = R'^2 \cos^{-1}(\frac{d'}{R'}) - d'\sqrt{R'^2 - d'^2} \quad (9)$$

twice, one for each half of the "lens." Noting that the heights of the two segment triangles are

$$d_1 = x = \frac{d^2 - r^2 - R^2}{2d} \quad (10)$$

$$d_2 = d - x = \frac{d^2 + r^2 - R^2}{2d}$$

The result is

$$A = A(R, d_1) + A(r, d_2) \quad (11)$$

$$= r^2 \cos^{-1}(\frac{d^2 + r^2 - R^2}{2dr}) - R^2 \cos^{-1}(\frac{d^2 - r^2 + R^2}{2dR}) \quad (12)$$

$$- \frac{1}{2}\sqrt{(d + r - R)(-d + r + R)(d - r + R)(d + r + R)}$$

Now we calculate the area fig 3.

***3.a***

If a, b, c>2R

$$s_h = s_\Delta - (s_1 + s_2 + s_3)$$

We know the total angular of triangular is:

$$\hat{\alpha} + \hat{\beta} + \hat{\zeta} = 1\hat{8}0$$

So

$$(s_1 + s_2 + s_3) = \pi R^2 \quad (13)$$

According to equation 1 we have

$$s_h = \sqrt{(a + b + c)(a - b + c)(b - c + a)(c - a + b)} - \pi R^2 \quad (14)$$

***3.b***

If a=b=c=2R

The area of hole calculates like section 3.a.

***3.c***

If a=b or a=c or b=c

The area of hole calculate like section 3.a .

***3.d***

$$s_h = s_\Delta - (s_1 + s_2 + s_3) + s_{\text{int}ersect}$$

According to equation 11 and By assume r=R we have

$$s_{\text{int}ersect} = R^2 \cos^{-1}(\frac{d^2}{2dR}) - R^2 \cos^{-1}(\frac{d^2}{2dR}) - \frac{1}{2}\sqrt{(-d + 2R)d^2(d + 2R)} \quad (15)$$

***3.e***

$$s_h = s_\Delta - (s_1 + s_2 + s_3) + \frac{1}{2}s_{\text{int}ersect1} + \frac{1}{2}s_{\text{int}ersect2} \quad (16)$$

$$s_{\text{int}ersect1} = 2R^2 \cos^{-1}(\frac{d_1^2}{2Rd_1}) - \frac{1}{2}\sqrt{(-d_1 + 2R)d_1^2(d_1 + 2R)}$$

$$s_{\text{int}ersect2} = 2R^2 \cos^{-1}(\frac{d_2^2}{2Rd_2}) - \frac{1}{2}\sqrt{(-d_2 + 2R)d_2^2(d_2 + 2R)}$$

***3.f***

It is obviously that there is not hole.

***3.g***

It calculates like 1.d .

***3.h***

It calculates like 1 .a

***3.i***

In this state we have three intersections region so we have (17)

$$S_h = S_\Delta - (S_1 + S_2 + S_3) + \frac{1}{2} S_{\text{int er sec t1}} + \frac{1}{2} S_{\text{int er sec t2}} + \frac{1}{2} S_{\text{int er sec t3}}$$

$$S_{\text{int er sec t1}} = 2R^2 \cos^{-1}\left(\frac{d_1^2}{2Rd_1}\right) - \frac{1}{2}\sqrt{(-d_1+2R)d_1^2(d_1+2R)}$$

$$S_{\text{int er sec t2}} = 2R^2 \cos^{-1}\left(\frac{d_2^2}{2Rd_2}\right) - \frac{1}{2}\sqrt{(-d_2+2R)d_2^2(d_2+2R)}$$

$$S_{\text{int er sec t3}} = 2R^2 \cos^{-1}\left(\frac{d_3^2}{2Rd_3}\right) - \frac{1}{2}\sqrt{(-d_3+2R)d_3^2(d_3+2R)}$$

In all of states if $S_h > 0$ then we conclude we have a hole.

## 4. Destination selection

In our proposed diagram (HSTT) After deciding the existence of a coverage hole and its size, a stationary node needs to decide the number of mobile nodes and the target locations of these mobile nodes to heal its holes .We want to aim maximum coverage so after calculate the area of hole we choice one of the circumcircle or incircle type. If the area is less than mobile sensor sensing region we use the circumcircle center for target location, if area of hole is larger than sensor sensing reign we use the incircle center for target location to aim maximum coverage.

The circumcenter of a triangle can be found as the intersection of the three perpendicular bisectors. (A perpendicular bisector is a line that forms a right angle with one of the triangle's sides and intersects that side at its midpoint.) This is because the circumcenter is equidistant from any pair of the triangle's points, and all points on the perpendicular bisectors are equidistant from those points of the triangle. It shows in Fig 6.

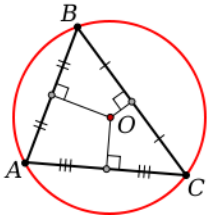



Fig.6 Construction of the circumcircle (red) and the circumcenter (red dot).

The incircle is the inscribed circle of a triangle $\Delta ABC$ i.e., the unique circle that is tangent to each of the triangle's three sides. The center $I$ of the incircle is called the incenter, and the radius $r$ of the circle is called the in radius. The incenter is the point of concurrence of the triangle's angle bisectors. In addition, the points $M_A$, $M_B$ and $M_C$ of intersection of the incircle with the sides of $\Delta ABC$ are the polygon vertices of the pedal triangle taking the incenter as the pedal point (c.f. tangential triangle). This triangle is called the contact triangle. It shows in fig 7.

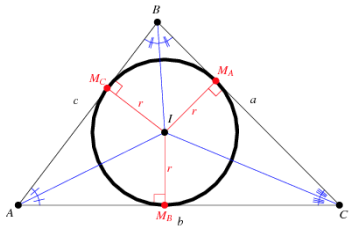



Fig.7 Construction of the incircle.

## 5. Conclusions

In this paper, we proposed a triangular oriented diagram (HSTT) that detect of hole and calculate its size. Also for maximum coverage we determine the target localization for mobile sensor for healing of coverage hole. All of them cause to along wireless network life time and optimization. In proposed diagram we use circumcircle and incircle to aim this goal. Our diagram compare with Voronoi diagram have some advantages such as

1. It is simple for construction.

2. It has lower than Voronoi diagram calculation for construction.

3. We get exact area of hole no estimation of it.

## 6. Future Work

We can emerge some adjacent triangular in our proposed diagram for use the low number of mobile sensor to healing coverage hole. Also we can use the hierarchical method according to size of hole to get maximum coverage by minimum mobile sensors. at the future we work to aim this goal.

## References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, “Wireless sensor networks: a survey”, Computer Networks 39 (4) (2002) 393–422.

[2] C.-F. Huang, Y.-C. Tseng, “A survey of solutions to the coverage problems in wireless sensor networks”, Journal of Internet Technology 6 (1) (2005) 1–8.

[3] M. Cardei, J. Wu, “Energy-efficient coverage problems in wireless ad hoc sensor networks”, Computer Communications 29 (4) (2006) 413–420.

[4] H. Zhang, J. Hou, “On deriving the upper bound of a-lifetime for large sensor networks”, in: ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), 2004, pp. 121–132.

[5] S. Kumar, T.H. Lai, J. Balogh, “On k-coverage in a mostly sleeping sensor etwork”, in: ACM International Conference on Mobile Computing and etworking (Mobicom), 2004, pp. 114–158.

[6] G.T. Sibley, M.H. Rahimi, G.S. Sukhatme, “Robomote: a tiny mobile robot latform for large-scale ad-hoc sensor networks”, in: IEEE International onference on Robotics and Automation, 2002, pp. 1143–1148.

[7] Y.-C. Tseng, Y.-C. Wang, K.-Y. Cheng, Y.-Y. Hsieh, “iMouse: an integrated mobile”.

[8] A. Ghosh, “Estimating coverage holes and enhancing coverage in mixed sensor etworks”, in: IEEE International Conference on Local Computer Networks, 2004, pp. 68–76.

[9] G. Wang, G. Cao, P. Berman, T.F.L. Porta, “Bidding protocols for deploying mobile sensors”, IEEE Transactions on Mobile Computing 6 (5) (2007) 515–528.

[10] F. Aurenhammer, Voronoi diagrams ,”a survey of a fundamental geometric data structure”, ACM Computing Surveys 23 (4) (1991) 345–406.

[11] G. Wang, G. Cao, T. LaPorta, “Proxy-based sensor deployment for mobile sensor networks”, in: IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS), 2004, pp. 493–502.

[12] Bang Wang, Hock Beng Lim, Di Ma, "A survey of movement strategies for improving network coverage in wireless sensor networks", Computer Communications 32 (2009) 1427–1436, Elsevier.

# Evolutionary Modular Neural Network Approach for Breast Cancer Diagnosis

**Bipul Pandey[1], Tarun Jain[2], Vishal Kothari[3] and Tarush Grover[4]**

**[1] Tata Consultancy Services Limited**
**New Delhi, Delhi, India**

**[2] Tata Consultancy Services Limited**
**New Delhi, Delhi, India**

**[3] Tata Consultancy Services Limited**
**New Delhi, Delhi, India**

**[4] Tata Consultancy Services Limited**
**New Delhi, Delhi, India**

**Abstract**

Knowledge Discovery paradigms especially Soft Computing techniques like Artificial Neural Networks have been at the fore front of research aimed at solving the problem areas involved in many diverse fields of application. Automated diagnosis of deadly diseases is one of such fields that have seen much effort from researchers in the last few years. One area where this effort has been most felt is the diagnosis of breast cancer in women. However, development of a computationally efficient, detection-wise effective and robust framework for the diagnosis of breast cancer has still not materialized. The major problem here is the presence of a number of decision variables involved that makes this problem of diagnosis much more complex and intricate. This makes it difficult to be tackled by traditional computing paradigms efficiently. In this paper, we explain how the paradigms of modularity and optimization using evolutionary technique could be used to solve the aforesaid problem with significant success. Here, to take benefit of modularity, we make of use modular neural network instead of the traditional monolithic neural network for the recognition of input vectors implying breast cancer. Also, to make the architecture more optimal, we make use of genetic algorithms to achieve optimal connections (weights) among the neurons in each of the individual experts of the modular neural network. Experimental results show that the proposed approach has been significantly successful in dealing with aforesaid problem of breast cancer diagnosis with a training accuracy of 95.97% and testing accuracy of 96.5%. That is well above what shown by traditional approaches as described later on.

***Keywords:*** *Breast Cancer Diagnosis, Genetic Algorithm, Modularity, Artificial Neural Network, Hybrid Computing.*

## 1. Introduction

Need for automated diagnosis of diseases has been closely felt in the last few decades. This need has been most acute in case of deadly diseases like cancer where early detection leads to much higher chances of successful treatment and recovery of the patient with considerable savings of financial resources. An example is the case of Breast Cancer in women. Breast is an extremely dangerous disease for the womenfolk. It has proved to be one of the biggest causes of mortality among women in the last many years. In USA, it is considered to be second leading cause of mortality among women and the most common cause of mortality in the age group 40 to 55 years among women [1]. In fact, in every thirteen minutes four American women develop this disease and one woman suffers death due to this disease [2]. Early detection is considered the best key to save the patient from the mortality due to this disease [3]. Because of such a scenario, researchers have focused their efforts for creating a diagnostic system for breast cancer.

Traditionally, diagnosis of diseases is done based on a number of tests done on the patient. The results of these tests are used by the medical practitioner to predict the presence/absence of a disease. However in case of many dangerous diseases, this task becomes rather myriad because the plethora of these tests not only confuses the medical practitioner, but also gives conflicting and non-conclusive results many a times. Hence, comes the need for an intelligent, automated

diagnostic system for diseases. As this problem can be computationally modeled as a problem of retrieving relevant information from a plethora of reports or tests, the development of a computationally efficient and detection-wise effective system for disease diagnosis can be seen be an issue from the field of Knowledge Discovery from Data (or KDD). KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [4]. Being cross-dimensional, Knowledge Discovery from data uses algorithms and techniques from a vast array fields like Soft Computing, Pattern Recognition, Machine Learning statistics, AI, Natural Language Processing etc. [5]. This field of Knowledge Discovery from data sources has seen a massive upsurge of interest from researchers for last few years. This is due to an increasing usage of information retrieved from data sources towards the development and functioning of intelligent systems in diverse applications along with an increasing capacity to capture and store very large amount of data. Also, with ever increasing demand for valuable information from seemingly mundane sources, researchers are concentrating on the task of developing novel methods and applications for the efficient discovery and retrieval of such information. However, the complex nature of the data and massive amount of inputs makes designing such methods an extremely difficult task [6].

As explained earlier, disease diagnostic via computational methods can be considered to be an application area of Knowledge Discovery from Data. However in addition to deal with the usual complexity of such tasks, here an added problem is the presence of a very large number of decision dimensions that used to classify the input vectors properly. In this paper, we have tried to deal with this task by proposing a hybrid framework. This framework combines the twin paradigms of modularity and genetic algorithm to lead a more optimal solution to the problem at hand. In place of a traditional monolithic neural classifier, we use a modular neural network that combines a number of individual experts to act independently upon the input to give individual outputs which are then combined by an integrator. In order to have an optimal architecture for each of the individual experts, we make use of genetic algorithm to give an optimal set of connections among the neurons of each of the individual experts involved. The breast cancer diagnosis system so developed consists of a Modular Neural Network for classifying the input data vectors as cancerous or non-cancerous. This classifier consists of six individual neural network experts. These individual experts are Feed-forward neural networks with single hidden layers. The integrator used to combine their outputs is the fuzzy C-means Integrator. Each of the individual experts is trained by using genetic algorithm with the training data set. After training, the system so obtained is tested on the testing data set to classify and diagnose the input data vectors as either belonging to a patient with breast cancer or a non-cancerous patient.

## 2.The State of Art

Today, there are a number of screening techniques being used for the detection of breast cancer. A few are: positron emission tomography (PET), magnetic resonance imaging (MRI), CT Scan, X-ray, ultra sound, photo-acoustic imaging, tomography, diffuse optical tomography, elastography, electrical impedance tomography, opto-acoustic imaging, ophthalmology, mammogram etc. Though, all of these have their own advantage mammogram is the most popularly used technique and is considered the most reliable [7]. But, even this technique suffers from some serious limitations. Up to 30% of the breast grazes couldn't be spotted in mammogram during screening. Also, images on mammogram could lead to not required biopsies [8].This absence of any fully effective, efficient method of breast cancer diagnosis has led to a spurt of efforts by researchers in the field of KDD towards developing an automated computational system for breast cancer diagnosis.

Soft Computing paradigms have been a valuable source of methods and techniques to be used for the task of discovering, capturing and retrieving knowledge and relevant information from data sources [9]. Several major soft computing paradigms including artificial neural network, fuzzy logic, evolutionary algorithms have found themselves being applied to the problem of KDD in general [1]. These paradigms have also found increasing usage for disease diagnostics especially breast cancer diagnosis. Early works focused on using simple feed-forward neural networks trained with back propagation algorithm for breast cancer diagnosis with screening techniques like mammography [10, 11]. Even with these primary investigations, the accuracy rate achieved was very significant as compared to other non-computational, human-intensive diagnostics. Such early successes prompted further investigation into the problem using other neural classifiers and techniques. In [12], Support Vector Machine (SVM) is successfully employed as a Classifier for diagnosing breast cancer. Principal Component Analysis (PCA) is used to extract

relevant knowledge in the form of features from ultrasound images of the patients. These relevant, independent features so obtained are then used as input to the SVM classifier to label each input as either cancerous or non-cancerous. Further research has focused on employing SVM based classifier along with feature selection technique for more accurate classification. Feature selection has been employed in order to reduce the number of inputs in order to decrease the complexity involved and increase computational efficiency [13]. Clustering has also been attempted on the problem. In [14], Self Organizing Map (SOM) technique has been used. Here, the analog video signal in sonography is used to obtain a digitized sonographic image. On this image, the SOM model is applied which uses 24 autocorrelation texture features for the classification task. Probabilistic Neural Network and General Regression Neural Network have also been applied on the said task [15]. The results so obtained show that General Regression Neural Network have been the most success in accurately identifying the nature of the input (cancerous or non-cancerous) as compared to other traditional neural classifiers used like Radial Basis Network (RBN) and Multi-Layer Perceptron (MLP). Some other researchers have used traditional data mining techniques like association rules along with artificial neural network to deal with the problem [16].

Though the use of these traditional monolithic neural models have been significantly successful for dealing with diagnostic task at hand, but further development of computationally more efficient and more accurate systems using neural network classifiers has suffered because of two major issues. The first is the problem of dimensionality. The problem of breast cancer diagnosis involves a large number of dimensions or attributes on which the classification is done and class labels decided. To deal with a large number of dimensions, Neural Network classifiers have to have a large number of neurons leading to a much more complex network. But such complex networks have lower performance [17]. Also, training becomes cumbersome and couldn't be done properly. The solution to this concern is the introduction of modularity. This is done by modular neural network which has a number of experts (neural networks) in contrast of the traditional monolithic neural network. This approach has been tested on the problem of breast cancer diagnosis. In [18], features selected from stepwise LDA have been classified by a modular neural network for the task of diagnosis. In several other efforts also, modularity approach has been found to be successful in providing a more efficient way for diagnosis of breast cancer [19, 20]. This paradigm of modular neural network has also been tested on similar problem of large decision dimensions like Biometrics, Financial Prediction [21, 22, 23, 24] and has been found to be much better as compared to other approaches involving traditional monolithic neural models.

The second concern is regarding the architecture of the neural network classifier used. The task of determining the architecture is human-intensive, hence prone to be sub-optimal. Researchers have tried to rectify this trouble by using evolutionary algorithms like genetic algorithm, particle swarm optimization etc. for getting an optimal topology (number of neurons, layers etc.) and optimal connections among neurons (i.e. evolutionary training). In [25], the problem of multi-modal biometrics is to be solved. Here, genetic algorithm has been used to optimize the modular neural network being used for the recognition task at hand. In [26], an evolutionary programming algorithm has been used to optimize both the connections and topology of the feed-forward neural network classifier being used. The optimized classifier thus used is then applied to the task of breast cancer giving fairly good results.

## 3.Methodology

In this paper, we have used the paradigms of modular neural network and genetic algorithm to deal with the twin concerns of dimensionality and sub-optimality of architecture. For the task of classifying the input data vectors, we have used a modular neural network (MNN) instead of a monolithic neural classifier. The MNN used comprises of six individual experts which independently work on the input vectors to produce their own output. These outputs are combined by an integrator (here, a fuzzy C-means integrator). In addition to it, genetic algorithm is used to obtain an optimal set of connections among the neurons in each of the individual experts involved by training each of the experts of the MNN.

### 3.1 Modular Neural Network

A Computational system that has two or more sub-systems that can work upon same or different inputs independently is said to show modularity. As such modular neural network are said to be those that comprise of two or more individual neural modules that can independently act on the inputs to produce output. This "Divide and Conquer" approach imbibes a number of advantages to such a neural network. These include complexity reduction in model,

scalability, flexibility in design and implementation, robustness, and computational efficiency [6]. These properties make modeling of problems with a large number of dimensions very efficient and easy while using modularity.

The framework of the proposed approach is as shown in figure 1.

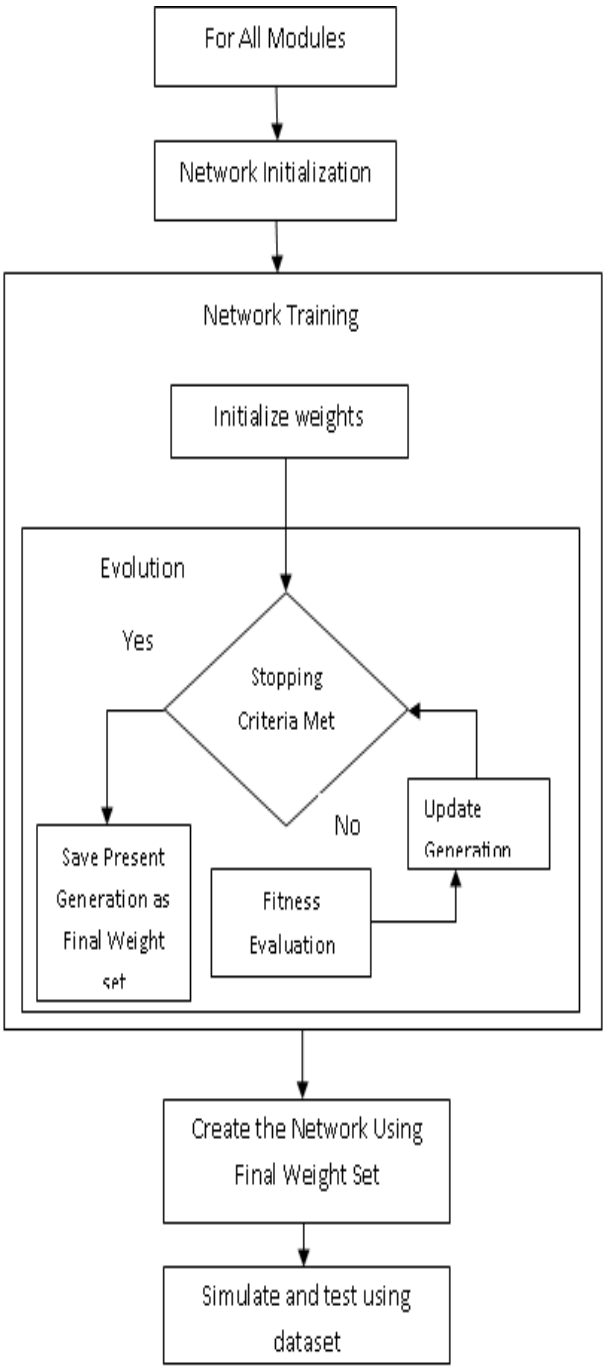



**Fig 1. The Framework of the Proposed Approach**

## 3.2 Genetic Algorithm

Genetic Algorithms are the most widely used among the evolutionary techniques for the task of optimization. They have been successfully used to give optimal architecture for the experts of a modular neural network. Therefore using the twin paradigms of modularity and genetic algorithms are considered to lead to far more efficient and successful solutions as compared to traditional monolithic modular neural networks [25, 26, 27]. Being an evolutionary technique, a Genetic Algorithm can be defined as a search and optimization heuristics that follows the natural process of evolution. The defining features that separate a GA from other evolutionary techniques are: Population of Chromosomes, selection according to fitness, crossover to produce new offspring and random mutation of new offspring [28]. Chromosomes are strings that encode prospective solutions for the problem being tackled. A population of these chromosomes is used for evolution. This evolution occurs over a number of generations. After each generation, the suitability (fitness) of each solution for the problem is checked using an objective function called fitness function according to which selection of solution set is done. These selected candidate solutions are then acted upon by genetic operators to produce the population set for the next generation. A few of these operators are: crossover, mutation, elite, add attribute, delete attribute, mutate number of neurons, and repair. Crossover uses two candidate solutions from the solution set and exchanges their data over a single or several crossover points to produce two candidate solutions. Mutation uses single candidate solutions from the solution set and changes a single data in the chromosome to give a new prospective solution. Mutation is used to give diversity to the solution. Elite takes a few of the fittest candidate solutions from the solution set and transfer them to the population of the next generation. It is used to preserve some local minima points in case one of them may be the global minima and hence, the most optimal solution to the problem. This process takes place until either the fitness threshold is achieved or number of generations is exhausted.

The first step here is to collect the relevant data for breast cancer from patients or subjects. The framework is to be used to classify and recognize which data set is cancerous or non-cancerous based on cell descriptions gathered by FNA image test. The data set used here is the breast cancer data from the UCI Machine Learning Repository for this purpose

(Wolberg, Mangasarian and Aha, 1992) [29]. The data set is then used to obtain training set and testing set. 70% of the entire data set goes into the training set while 30% goes into the testing set. Each data set comprises of data vectors with 30 decision variables and a class variable. The training data set is to be used for the supervised learning of each individual expert of the modular neural network. Now, the modular neural network is initialized. The modular neural network comprises of six individual neural modules. Each of these is a Feed-forward Neural Network. Each of the experts has one input layer with 30 neurons and one hidden layer with 30 neurons. The output layer has two nodes. Now, each of the experts is trained. Here, we have relied on Genetic Algorithm for achieving optimized connections among the neurons. This is done by using GA for training each of the experts. The genetic operators to be used in the approach are: Crossover (70%), Mutation (20%) and Elite (10%). For each expert, the weights are randomly initialized in the set [1,-1] equal to the number of the connections among the neurons of the expert. 30 such sets are taken as the initial population for the GA. The maximum number of generations is taken as 100. The maximum number of generations is also the stopping criteria for the Genetic Algorithm. After the first generation, the fitness of each of its member is evaluated.

The fitness function used here is the Root Mean Square Error (RMSE):

$$RMSE = (\sum(f(x_i) - y_i)^2 / 2)^{1/2} \quad (1)$$

Where $f(x_i)$ is the target and '$y_i$' is the actual value and 'n' is the number of patterns used for training.

After this, the population is acted upon by crossover, mutation and elite to produce the population for the next generation. This process is carried on till the maximum number of generation is exhausted. After this, the set with the least RMSE is taken as the set of weights for the individual experts of the modular neural network. After using GA to train each of the experts in the way described above, the testing data set is used to test the approach.

## 4. Experimental Results

The objective is to apply and check the performance of the proposed approach (using training and testing accuracy) over the breast cancer diagnosis problem. The proposed approach is used to classify the given data vectors of the subjects as either cancerous or non-cancerous. The dataset used is the breast cancer data from the UCI Machine Learning Repository for this purpose (Wolberg, Mangasarian and Aha, 1992) [29]. The data set comprises of data vectors from 569 patients out of which 212 patients have breast cancer. Each data vector of the data set comprises of 30 decision attributes and a single class attributes. Attributes in the data set include radius mean of distances from center to points on the perimeter, texture means standard deviation of gray-scale values, smoothness means local variation in radius lengths, perimeter, area, smoothness (local variation in radius lengths) , compactness (perimeter2 / area - 1.0), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry and fractal dimension (coastline approximation - 1).These are measured for a total of 3 cells. The data set has been initially divided into a training set and a testing set. The training set comprises of 398 vectors i.e. about 70% of the data set. The testing data set comprises of the rest 30% of the data set.

Matlab is used as the implementation platform. The Modular Neural Network has been coded on the Matlab while Genetic Algorithm of the GA toolbox is used in the implementation.

Initially, each of the individual neural modules of the MNN is initialized and trained with Genetic Algorithm using the training set. After this, the approach is tested using the testing data set. This process is repeated fifteen times. Then the mean training accuracy and the mean testing accuracy is calculated by calculating the average number of correctly identified and incorrectly identified data vectors. The results obtained are as listed in Table 1.

**Table 1. Experimental Results Obtained from the Proposed Approach**

| S. No. | Property | Value |
|---|---|---|
| 1. | Mean Training Accuracy | 95.97% |
| 2. | Mean Testing Accuracy | 96.5 % |
| 4. | Mean Correctly Identified Instances (Training) | 382 |
| 5. | Mean Incorrectly Identified Instances (Training) | 16 |
| 6. | Mean Correctly Identified Instances (Testing) | 165 |
| 7. | Mean Incorrectly Identified Instances (Testing) | 6 |

For comparison, we have also implemented and used four other widely used approaches for the same task in Matlab. These were: Multi-Layer Perceptron

(MLP) with BPA training, Fixed Architecture Evolutionary ANN, Variable Architecture Evolutionary ANN, and Modular Neural Network. The comparative results of these approaches along with that of the proposed approach could be seen in Table 2.

**Table 2. Comparison of Experimental Results Obtained from Various Approaches**

| S. No. | Algorithm | Training Accuracy | Testing Accuracy |
|---|---|---|---|
| 1. | Proposed Approach | 95.97 % | 96.5 % |
| 2. | MLP with BPA | 97.10% | 94.52% |
| 3. | Fixed Architecture Evolutionary ANN | 94.00% | 95.27% |
| 4. | Variable Architecture Evolutionary ANN | 97.16% | 95.00% |
| 5. | Modular Neural Network | 97.54% | 95.60% |

## 5. Conclusion

The experimental results show that the proposed approach shows a very high training and testing accuracy for breast cancer diagnosis. The accuracy achieved for both training and testing is much better than that of the four other popular approaches used here. This shows that the proposed approach could be used to give better solutions to complex problems where we have deal with the problem of dimensionality.

Hence, the proposed approach could be used for other such complex problems like Biometrics, Robot Coordination etc. Also, other optimization techniques could be used in place of Genetic Algorithm like Particle Swarm Optimization, Ant Colony Optimization etc. All this is planned to be done in future.

## References

[1] www.breastcancer.org.
[2] V.F. Andolina, S.L. Lille, and K.M. Willison, Mammographic Imaging: A Practical Guide. New York, NY: Lippincott Williams & Wilkins, 1992.
[3] S. S. Basha, and K. S. Prasad, "Automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy c – means clustering', Journal of Theoretical and Applied Information Technology, pp: 704-709.
[4] U. M. Fayyad, G. P. Shapiro, and P. Smyth, From Data Mining to Knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data Mining, AAAI Press, 1996.
[5] P. N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Pearson Addison Wesley, 2005.
[6] M. H. Dunham, Data Mining Introductory and Advanced Topics, Prentice Hall, 2003.
[7] L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Detection and classification of mammographic calcifications," in State of the Art in Digital Mammographic Image Analysis, World Scientific, 1994.
[8] B. Adam, S. J. Nover, W. Anjum, H. Yegingil, W. Y. Shih, W. Shih, and A. D. Brooks, "Computer applications for early detection and staging of cancer", International Journal of Biomedical Imaging, 2009, pp. 1-14.
[9] J. Han, and M. Kamber, Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann, 2005.
[10] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C.E. Metz, "Artificial neural networks in mammography: application to decision making in diagnosis of breast cancer", Radiology, 1993.
[11] P. Wilding, M. A. Morganb, A. E. Grygotisa, M. A. Shoffnera, and E. F. Rosato, "Application of backpropagation neural networks to diagnosis of breast and ovarian cancer", Cancer Letters Vol. 77, No. 2-3, 1994, pp. 145-153.
[12] Y. L. Huang, D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu, and W. K. Moon, "Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound", Ultrasound Obstet Gynecol, 32(4), 2008, pp. 565-572.
[13] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis", Expert Systems with Applications, Vol. 36, No. 2, Part 2, 2009, pp. 3240-3247.
[14] D. Chen, R. Chang, and Yu-Len Huang, "Breast cancer diagnosis using self-organizing map for sonography", Ultrasound in Medicine & Biology Vol. 26, No. 3, 2000, pp. 405-411.
[15] T. Kiyan, and T. Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks", Journal of Electrical & Electronics Engineering, Vol. 4, No. 2, 2004, pp. 1149-1153.
[16] M. Karabataka, and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications,Vol. 36, No. 2, Part 2, 2009, pp. 3465-3469.

[17] F. Azam, "Biologically Inspired Modular Neural Networks", PhD Dissertation, Virginia Tech. 2000.
[18] Y. L. Joseph, G. Marios, K. M. Mia, and L. J. Jonathan, "Computer-aided classification of breast microcalcification clusters: Merging of features from image processing and radiologists", in Proceedings of SPIE, 2003, Vol. 5032.
[19] E. D. Ubeyli, "A Mixture of Experts Network Structure for Breast Cancer Diagnosis", Journal of Medical Systems, Vol. 29, No. 5, 2005.
[20] F. Schnorrenber , N Tsopafsoulis, S. Iollios, M. Vossiliou, A. Adamou, and K. Iyriacoui, "Improved Detection of Breast Cancer Nuclei Using Modular Neural Networks", IEEE Engineering In Medicine And Biology, 2000.
[21] F. Gaxiola, P. Melin, and M. López, "Modular Neural Networks for Person Recognition Using the Contour Segmentation of the Human Iris Biometric Measurement", Soft Computing for Recognition Based on Biometrics, Studies in Computational Intelligence, Vol. 312, 2010, pp. 137-153.
[22] M. Serrano, and P. Melin, "A Modular Neural Network with Fuzzy Response Integration for Person Identification Using Biometric Measures", Evolutionary Design of Intelligent Systems in Modeling, Simulation and Control, Studies in Computational Intelligence, Vol. 257, 2009, pp. 159-183.
[23] R. Kala, H. Vazirani, A. Shukla, and R. Tiwari, "Fusion of Speech and Face by Enhanced Modular Neural Network", Information Systems, Technology and Management, Communications in Computer and Information Science, Vol. 54, No. 6, 2010, pp. 363-372.
[24] N. Gradojevic, R. Gencay, and D. Kukolj, "Option Pricing With Modular Neural Networks", IEEE Transactions on Neural Networks, Vol. 20, No. 4, 2009, pp. 626 – 637.
[25] P. Melin, A. Mancilla, M. Lopez, and O. Mendoza, "A hybrid modular neural network architecture with fuzzy Sugeno integration for time series forecasting", Applied Soft Computing, Soft Computing for Time Series Prediction, Vol. 7, No. 4, 2007, pp. 1217-1226.
[26] J. M. Villegas, A. Mancilla, and P. Melin, "Optimization of Modular Neural Network, Using Genetic Algorithms: The Case of Face and Voice Recognition" Soft Computing for Hybrid Intelligent Systems, Studies in Computational Intelligence, Vol. 154, 2008, pp. 151-169.
[27] X. Yao, "Evolving Artificial Neural Networks", in Proceedings of the IEEE, 1999, Vol. 87(9), pp: 1423-1447.
[28] M. Mitchell, An Introduction to Genetic Algorithms, Cambridge, Massachusetts: MIT Press, 1999.
[29] W. H. Wolberg, O. L. Mangasarian, and D. W. Aha, UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html] University of Wisconsin Hospitals, 1992.

**Bipul Pandey** received his Bachelor (B.Tech) and Master degree (M.Tech) from Indian Institute of Information Technology and Management, Gwalior, India. He is currently employed as Assistant Systems Engineer at Tata Consultancy Limited, India. He has authored six research papers published in reputed international journals and conferences. His areas of research are soft computing, hybrid system design, Bioinformatics, Biomedicals and Biometrics. He focuses his research initiatives for presenting innovative technological solutions involving soft computing for connecting common problems to their more intelligent solutions.

**Tarun Jain** received his Bachelor Degree in Biotechnology from Thapar University, Patiala, India. He is currently Assistant Systems Engineer at Tata Consultancy Services, India.He is author of various International Journals in field of Biopharmaceuticals and Nanobiotechnology. Also, his research has been presented in various National and International Conferences. He focuses his interdisciplinary research on the field of Life science and Healthcare related issues.

**Vishal Kothari** received his Bachelors Degree From MIT, Ujjain, India. He received his Masters Degree from Indian Institute of Information Technology and Management, Gwalior, India. He is currently employed as Assistant Systems Engineer at Tata Consultancy Limited, India. He focuses his research initiative in the field of Bioinformatics and Soft Computing.

**Tarush Grover** received his Bachelor Degree from Thapar University, Patiala, India. He is currently Assistant Systems Engineer at Tata Consultancy Services, India. His areas of interest include Soft Computing and Biometrics.

# Numerical simulation of groundwater level in a fractured porous medium and sensitivity analysis of the hydrodynamic parameters using grid computing: application of the plain of Gondo (Burkina Faso)

**Wenddabo Olivier Sawadogo[1], Noureddine Alaa[2] and Blaise Somé[3]**
**'Laboratoire d'Analyse Numérique, d'Informatique et de Biomathématique, Universitéde Ouagadougou, 03 B.P. 7021 OUAGADOUGOU 03, Burkina Faso,**
**[2]Laboratoire de Mathématiques Appliquées et d'Informatique, B.P. 549, Av. Abdelkarim Elkhattabi, Guéliz Marrakech, Maroc,**
**[3]Laboratoire d'Analyse Numérique, d'Informatique et de Biomathématique, Universitéde Ouagadougou, 03 B.P. 7021 OUAGADOUGOU 03, Burkina Faso,**

## Abstract

The use of mathematical modeling as a tool for decision support is not common in Africa in solving development problems. In this article we talk about the numerical simulation of groundwater level of the plain of Gondo (Burkina Faso) and the sensitivity analysis of the hydrodynamic parameters. The domain has fractures which have hydraulic coefficients lower than those of the rock. Our contribution is to bring brief replies to the real problem posed in the thesis of Mr. KOUSSOUBE [1]. Namely that what causes the appearance of the piezometric level observed and impact of surface water on the piezometry. The mathematical model of the flow was solved by programming the finite element method on FreeFem++[2]. A local refinement of the mesh at fracture was used. We then conduct a sensitivity analysis to see which hydrodynamic parameters influences much of the solution. The method used for the sensitivity analysis is based on the calculation of the gradient by the adjoint equation and requires great computational power. To remedy this, we used a technique of distributed computing and we launched our application to the Moroccan grid (magrid). This allowed us to reduce the computation time. The results allowed to highlight the role of fractures and contributions of surface water on the evolution of the piezometric level of the plain of Gondo and identified the parameters that greatly influence the piezometric level.

**Keywords:** *fractured porous media, finite elements, hydrodynamic parameters, sensitivity analysis, distributed computing, grid computing*.

## 1 Introduction

The goal of this work is the mathematical modeling and numerical simulation of groundwater level of the plain of Gondo. The mathematical model was solved by the Galerkin finite element method. The programming was done in FreeFem++ [2].

We worked on a geological section of reference of the plain of Gondo. The geological section shows a fractured porous media [1] (see Figure 1). The usual models programmed with standard softwares asume that the medium is homogeneous and continuous [3]. Fractured aquifers cannot be modeled in a simple way. The flow in the fractures does not meet the laws that govern a continuous medium. The presence of fractures greatly influences the hydraulic conductivity. In most cases, the fractures have a hydraulic conductivity greater than that of the rock. In this case they constitute privileged channels for the water circulation. In our case, the fractures have hydraulic conductivities lower than the rock. There are several methods and models for treating fractures [4] [5]. To treat the presence of fractures, we assume that the fractures are sub-domains of a global domain as in [6]. We therefore assimilate them to porous media and we use finer meshes in fractures. Each domain was meshed separately and we have imposed on the nodes common border of two neighboring domains.

Our paper is organized as follows. In the first part, we present the problem to be solved. The second part is devoted to mathematical modeling and to the numerical solution of the model used. A third part is devoted to presenting the results of numerical simulations, followed finally by a fourth section on sensitivity analysis of hydrodynamic parameters, in wich we present the analysis method used and this execution on the grid computing.

## 2 Problem to be solved

### 2.1 Site description: geology and geometry of the site [1]

The plain of Gondo is vast; by lack of data on the whole plain, the domain used for the modeling is a geological section located on the edge of the basin of Gondo (Figure 1). On this zone, many investigations were realized within the framework of the thesis [1]: geophysics, piezometry, correlation of logs of drillings,

chemistry and isotopic geochemistry, observations of ground. This vertical section of selected reference extends from the village of Nomou located at the East in the crystalline base to the village of Yensélocated at the West in the sedimentary basin, that is to say a length of 25 km approximately. It has a height of 300 m. On the section of reference, five subverticale fractures of 500 m thickness affect the base and the sedimentary formations as well.

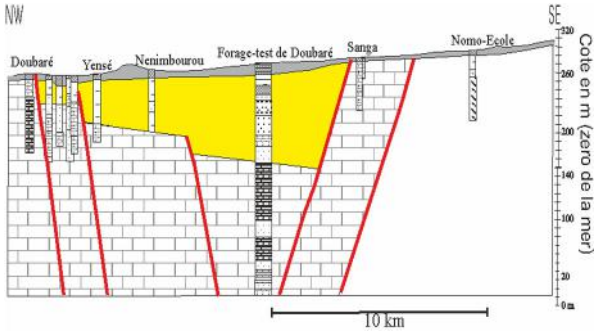



Figure 1: Geology and geometry of the site

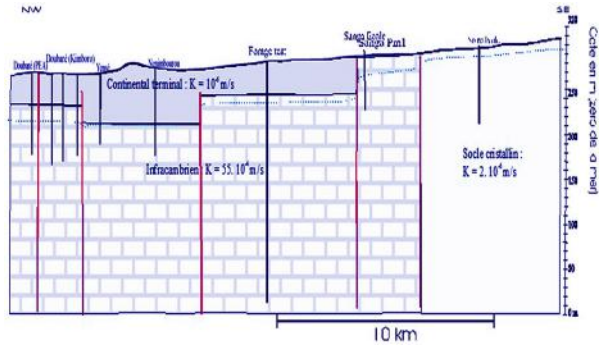



Figure 2: Simplified interpretative geological section Appearance of the piezometry: dotted curve in blue

## 2.2 Boundary conditions for the flow

The piezometry observed with the upstream (zone of base to the upstream) and with the downstream (sedimentary zone with the downstream of Doubaré) of the flow is prescribed as a limit with imposed potential. The lower limit of the field constitutes a limit with null flow. We must also take into account the recharge induced by rainfall in the upper part.

## 2.3 Responses required for flow modeling

The mathematical modeling and numerical simulation must bring answers to the following questions:

- What is the role of the fractures on the stair appearance observed of the piezometry (Figure 2)?
- What is the impact of the concentration of surface water on the evolution of the piezometry?

To provide some answers to these questions, we will do a simulation in steady state level of the water in two dates (1960 and 2000).

# 3 Mathematical model and numerical solution

## 3.1 Mathematical model

We consider the steady flow of an incompressible and monophasic fluid in a saturated porous medium Ω. The flow is governed by the law of conservation of mass and Darcy's law[3][7]:

$$\begin{array}{rll} div(u) = f & \text{in} & \Omega \\ u = -K\nabla p & \text{in} & \Omega \\ p = d & \text{on} & \Gamma^0 \\ u.\nu = -K\frac{\partial p}{\partial n} = g & \text{on} & \Gamma^1 \end{array} \qquad (1)$$

where u(x) is the Darcy velocity, p(x) is the hydraulic potential, f(x) the source term, K(x) the hydraulic conductivity, Ω is a bounded open of $\mathbb{R}^2$, d(x) is the Dirichlet boundary conditions and g(x) the Neumann boundary conditions.

Remark: in our case g(x) = 0 (no-flux), d(x) and K(x) are piecewise constant functions.

Subdivide our domain into several sub-domains Ωi,

(i = 1,2,.. . , 13) (see Figure 3). The

fractures are treated as porous media. We notice

Γi = Ωi ∩ Ωi+1, i = 1,2,... ,12

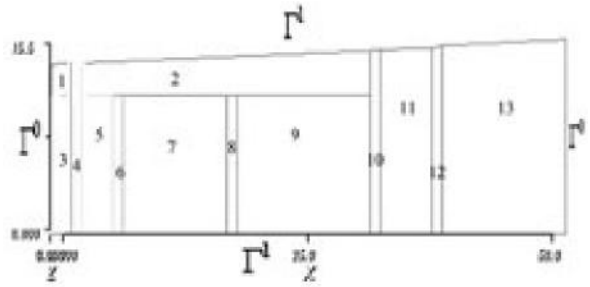


Figure 3: Subdivided domain

We introduce below the variational formulation of problem (1), which is equivalent to the following transmission equations [8]:

$$\begin{array}{rcl} div(u_i) = f_i & \text{in} & \Omega_i \\ u_i = -K_i \nabla p_i & \text{in} & \Omega_i \\ p_i = d_i & \text{on} & \Gamma_{Di} \\ u_i.\nu_i = 0 & \text{on} & \Gamma_{Ni} \\ p_i = p_{i+1} & \text{on} & \Gamma_i \\ u_i.\nu_i = u_{i+1}.\nu_{i+1} & \text{on} & \Gamma_i \end{array} \quad (2)$$

These equations reflect the exchange between the aquifers.
By eliminating u in equation (1), we have formally:

$$\begin{array}{rcl} -div(K\nabla p) = f & \text{in} & \Omega \\ p = d & \text{on} & \Gamma^0 \\ -K\frac{\partial p}{\partial n} = g & \text{on} & \Gamma^1 \end{array} \quad (3)$$

## 3.2 Problem solving

Let V be the space be defined by

$V = \{v \in H^1(\Omega), v = 0 \text{ on } \Gamma^0\}$.

Since $mes(\Gamma^0) > 0$, according to [9], we can choose

$$\|v\|_V = \left(\int_\Omega |\nabla v|^2\right)^{1/2} \text{ as norm on } V.$$

Let $v \in V$ be a test function. On multiplying (3) by v and integrating by parts, the variational formulation associated to the problem (3) is:

$$\begin{cases} \text{Find } p \in H^1(\Omega) \text{ such that } p = d \text{ on } \Gamma^0 \text{ and such that} \\ \int_\Omega K\nabla p.\nabla v = \int_\Omega fv - \int_{\Gamma^1} gv, \forall v \in V \end{cases} \quad (4)$$

We denote by $\gamma 0$ the trace operator. Let $r_d \in H^1(\Omega)$ such as $\gamma 0(r_d) = d$ and we denote $p_0 = p - r_d$. The variational formulation becomes:

$$\begin{cases} \text{Find } p_0 \in V \text{ such as} \\ \int_\Omega K\nabla p_0.\nabla v = \int_\Omega K\nabla r_d.\nabla v + \int_\Omega fv - \int_{\Gamma^1} gv, \forall v \in V \end{cases} \quad (5)$$

The function K(x) is constant in each domain Ωi, i = 1, 2, ... ,13. We have $K(x) = \sum_{i=1}^{13} K_i I_i(x)$ where $K_i > 0$ and where the function $I_i(x)$ is the characteristic functions of the domain Ωi for i = 1, 2, ... , 13.
We have then $0 < \min(Ki) < K(x) < \max(Ki)$. We assume that $g \in L^2(\Gamma^1)$ and $f \in L^2(\Omega)$.
Let the bilinear form $\alpha : V \text{x} V \to \mathbb{R}$ be defined by:

$$\alpha(p_0, v) = \int_\Omega K\nabla p_0.\nabla v$$

Let the linear form $L : V \to \mathbb{R}$ be defined by:

$$L(v) = \int_\Omega K\nabla r_d.\nabla v + \int_\Omega fv - \int_{\Gamma^1} gv$$

The space V is a Hilbert space for the Hilbertian norm $\|.\|_V$.
The bilinear form $\alpha$ is continuous, coercive and the linear form L is also continuous. Thus the theorem of Lax-Milgram [10] ensures the existence and uniqueness of a solution to the variational problem (5) and consequently the existence and uniqueness of a solution of (3).
Let $T_h$ be a triangulation of Ω. Let $P_1$ denote the space of continuous, piewise affine function in Ω i.e the space of continuous functions which are affine in x, y on each triangle of $T_h$. We pose $V_h = P_1 \cap V$. $V_h$ is a linear vector space of finite dimension. We denote N its dimension and $\phi_1, \ldots, \phi_N$ a basis. The approximated problem is:

$$\text{find } p_h \in V_h, \text{ such that } \alpha(p_h, v_h) = L(v_h) \text{ for all } v_h \in V_h \quad (6)$$

Let

$$p_h(x, y) = \sum_{i=1}^{N} p_i \phi_i(x, y)$$

and take $v_h = \phi_i$ for i = 1,..., N; equation (6) is equivalent to

$$\alpha(\sum_{j=1}^{N} p_j \phi_j, \phi_i) = L(\phi_i), \ i = 1, ..., N \quad (7)$$

This gives the system $Ax = b$, where:

$$A_{ij} = \int_\Omega K\nabla\phi_i.\nabla\phi_j = \sum_{T \in T_h} \int_T K\nabla\phi_i.\nabla\phi_j$$

$$b_i = \int_\Omega f\phi_i + \int_\Omega K\nabla r_d.\nabla\phi_i - \int_{\Gamma^1} g\phi_i$$

$$= \sum_{T \in T_h} \int_T f\phi_i + \sum_{T \in T_h} \int_T K\nabla r_d.\nabla\phi_i - \sum_{T \in T_h} \int_{T \cap \Gamma^1} g\phi_i$$

# 4 Results and discussions

## 4.1 Problem test

Before applying our code to our case, we compared our results to a reference case obtained by the method of domain decomposition [11]. The length is 2cm, the width is 1cm and the opening of the fracture is d = 0.01. Hydraulic conductivity at the fracture $K_f$ = 100 and that of the rock is 1. The results are shown in Figure 4.
**Remark:** We multiplied the size of the area by 10.

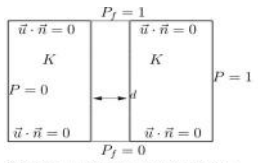



figure 2.1 : Domain with boundary conditions

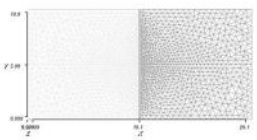


figure 2.2: Mesh

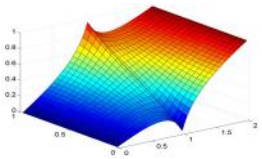


figure 2.3: Pressure obtained par domain decomposition method[11].

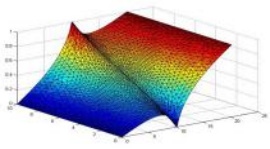


figure 2.4: Pressure obtained by local refinement of the mesh at the fracture (approach used in our case)

Figure 4: test case

## 4.2 Application to our problem

We used a triangular mesh. To have a high degree of accuracy we chose an average step of 50 m (horizontally) and 10 m (vertically) (see Figure 5). The data used for the realization of the simulations are from the thesis of Mr Youssouf Koussoubé[1]. We have
K3 = K5 = K7 = K9 = K11 and K1 = K2 because sub-domains 3, 5, 7, 9 and 11 are the same geological formation, and so are sub-domains 1 and 2. So we have a total of 8 hydrodynamic parameters.

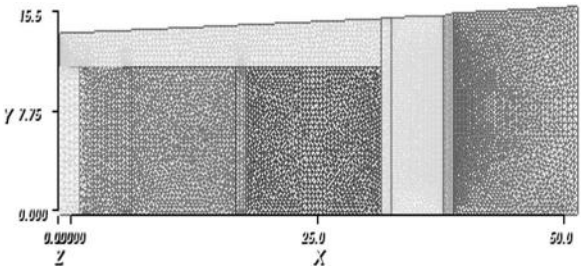



Figure 5: Mesh

### 4.2.1 Simulation 1: level of 1960

**Hydrodynamic parameters**
K1 = $10^{-6}$m/s, K2 = $55.10^{-4}$m/s, K3 = $10^{-6}$m/s,
K4 = $10^{-8}$m/s, K5 = $10^{-8}$m/s, K6 = $10^{-8}$m/s, K7 = $5.10^{-8}$m/s,
K8 = $2.10^{-6}$m/s.

**Boundary conditions and source term**
The potential imposed on the upstream is 222 m and 199 m on downstream. The source term is the groundwater recharge by rainfall that is 120 mm/year.

Figure 6 represents the result of this simulation by plotting the potential as a function of the space variables x and y.

### 4.2.2 Simulation 2: level of 2000

The hydrodynamic parameters values are the same than 1960. The potential imposed on the upstream is 287 m and 211 m on downstream. The source term is the groundwater recharge by rainfall that is 100 mm/year. The graphics of the Figure 7 represent the results of the simulation.

### 4.2.3 Role of fractures and surface water on the rise of piezometry

To see the impact of the fractures and surface water on the piezometry, we realised two simulations.

**Simulation 3 (Figure 8)**
The boundary conditions are those of 2000. In addition to the value of the source term of 2000, we added between fracture 4 and fracture 5, the contributions of surface water is 2600 million $m^3$ of water. We note a rise of piezometry compared to that of 2000 (see Figure 10).

**Simulation 4(Figure 9)**
In addition to the preceding test conditions, we modified the hydrodynamic properties of the fractures.
K1 = $10^{-6}$m/s, K2 = $55.10^{-4}$m/s, K3 = $10^{-6}$m/s, K4 = $10^{-3}$m/s,
K5 = $10^{-3}$m/s, K6 = $10^{-6}$m/s, K7 = $5.10^{-7}$m/s, K8 = $2.10^{-6}$m/s.
We note that the groundwater level has not increased despite the addition of the source term (see Figure 10). That shows the influence of the hydrodynamic properties of fractures on the piezometry.

### 4.2.4 Role of the fractures on the pace observed of piezometry

To see the role of the barriers on piezometry in stair observed we conducted two simulations by modifying the properties of the fractures. The other values are those of 1960.

**Simulation 5 (Figure 11)**
K1 = $10^{-6}$m/s, K2 = $55.10^{-4}$m/s, K3 = $10^{-3}$m/s, K4 = $10^{-3}$m/s,
K5 = $10^{-3}$m/s, K6 = $10^{-3}$m/s, K7 = $5.10^{-3}$m/s, K8 = $2.10^{-6}$m/s

**Simulation 6(figure12)**
K1 = $10^{-6}$m/s, K2 = $55.10^{-4}$m/s, K3 = $10^{-3}$m/s, K4 = $10^{-3}$m/s,
K5 = $10^{-3}$m/s, K6 = $10^{-8}$m/s, K7 = $5.10^{-8}$m/s, K8 = $2.10^{-6}$m/s.

Simulations 1 to 4 make it possible to conclude that the rise of the piezometry is the result of two factors: the contribution of surface water, between fractures 4 and 5, as well as the low permeability of the fractures which prevents the flow.
As for the stair appearance observed of the piezometry, seems to be the result of the hydrodynamic property of the fractures which cause jumps of pressures, i.e they cause discontinuities of the pressure. This is shown to us by the simulations 5 and 6.
For order to better see the impact of the hydrodynamic parameters on the piezometry, we will carry out an analysis of sensitivity.

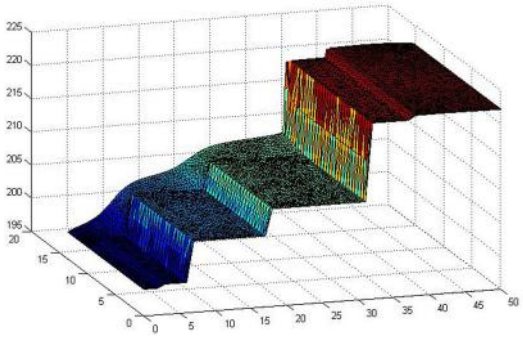


Figure 6: Simulation 1

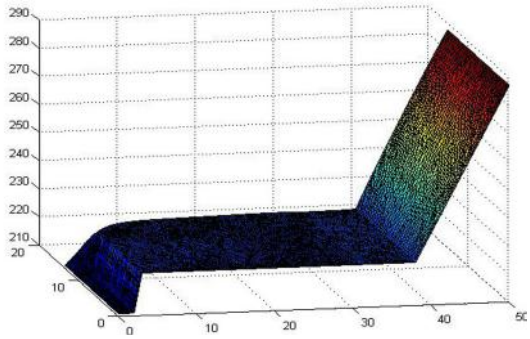


Figure 9: Simulation 4

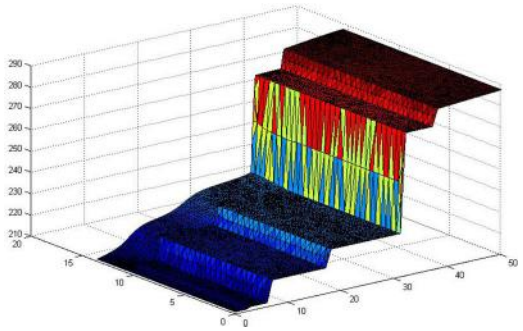


Figure 7: Simulation 2

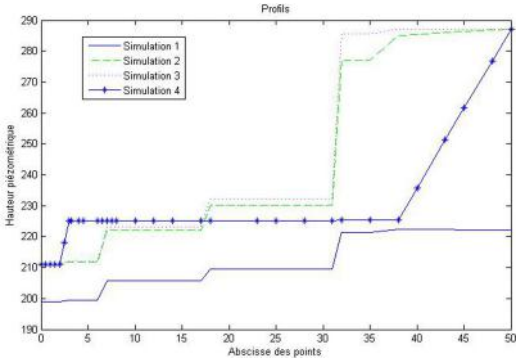



Figure 10: Profiles of simulations 1,2,3,4

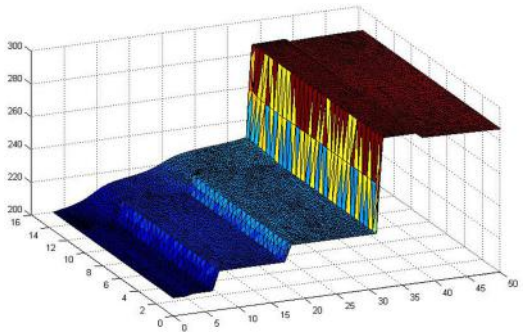


Figure 8: Simulation 3

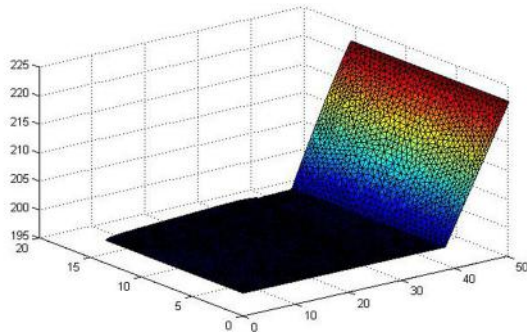


Figure 11: Simulation 5

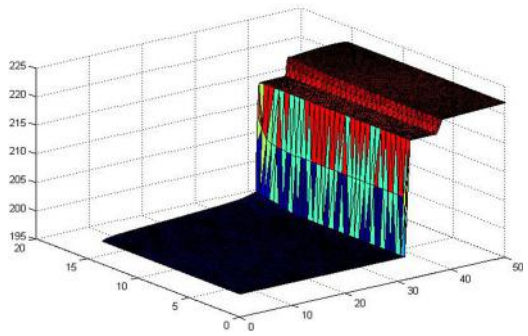


Figure 12: Simulation 6

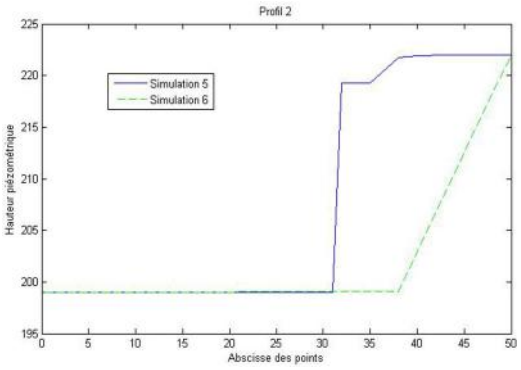



Figure 13: Profiles of simulation 5 and 6

# 5 Analysis of sensitivity of hydrodynamic parameters

## 5.1 Interest of the analysis of sensitivity

A mathematical model is a simplified or skewed representation, more or less realistic of the variable of state which it simulates. It is particularly the case for the models of flow, since one does not know exact equations governing the laws of their variables of state. In fact, to know uncertainty on the outputs of the model is essential. When a model is used, it is difficult to guess which parameters will have the most weight in the model, on which to pay more attention in term of precision, on which a disturbance would generate a consequent difference at exit. Given a model, one may ask what parameters must be estimated prioritarily, how a small change of control will impact the output. And even for economic reasons or practices, we may wish to consider the impact of the frequency of these observations on the results of estimating model parameters. Sensitivity analysis which is by definition the study of the impact of control variables on the output; it can provide some answers to these concerns. It is possible to group the methods of sensitivity analysis in three classes [12]: the methods of screening, which consist in a qualitative analysis of the sensitivity of the output variable to input variables, local methods analysis (based on the calculation of the derivative), which assess quantitatively the impact of a small variation around a given value of the inputs, and finally the methods of global sensitivity analysis (based on the analysis Statistics), interested in the variability of the model output in its entire range of variation. In this work we will conduct a sensitivity analysis to see the local impact of each control variable on the state variable.

## 5.2 Definition

A sensitivity analysis involves:

• a model

$$F(X,K) = 0 \qquad (8)$$

where $X$ is the state variable ; $K$ are the control variables of the model; $F$ a differential operator that is non-linear a priori, finite-dimensional.

Given $K$, we suppose that the system (8) has a unique solution.

• a response function G, function of (X, K) with scalar value, which expresses (in a certain way adapted to the situation) one or several outputs of the model of which one tries to assess the sensitivity.

It then seeks to determine the sensitivity of G with respect to K. Mathematically the sensitivity S of G over K is defined as the gradient of G with respect to K [13]:

$$S = \nabla_K G \qquad (9)$$

In our case the mathematical model is given by equation (3). The variable of state is the pressure p and the variables of control are given by the vector $K = (K_1, \ldots, K_8)$.

## 5.3 Definition of the function G

Let $p_r$ be a reference solution of equation (3) corresponding to a reference parameter vector $K_r$. We seek the impact of the disruption $K_r$ on $p_r$. For this we take the functional J defined below as the criterion.

$$J(K) = \|p(K) - p_r\|^2_{L^2(\Omega)} \qquad (10)$$

By definition calculate the sensitivity S is to calculate the gradient of J at point $K(\nabla_K J)$.
To do this we will go through the adjoint state method.

## 5.4 Calculating the gradient by the adjoint equation

We have:

$$J(K) = \int_\Omega (p(K) - p_r, p(K) - p_r) d\Omega$$

**Calculation of the directional derivative:**

By definition, we have:

$$J(K)[k] = \lim_{\beta \to 0} \frac{J(K + \beta k) - J(K)}{\beta}$$

After calculation, we have:

$$J(K + \alpha k) - J(K) =$$

$$\int_\Omega (p(K + \beta k) + p(K) - 2p_r).(p(K + \beta k) - p(K)$$

Dividing by β and taking the limit, we have:

$$J(K)[k] = \int_\Omega 2(p(K) - p_d).\hat{p} d\Omega$$

where

$$\hat{p} = \lim_{\beta \to 0} \frac{p(K + \beta k) - p(K)}{\beta}$$

**Tangent linear model:**
As p(K + βk) and p(K) are solutions of the equation (3) we show that $\hat{p}$ is solution of

$$\begin{array}{rll} -div(K\nabla\hat{p}) = div(k\nabla p) & \text{in} & \Omega \\ \hat{p}(x) = 0 & \text{on} & \Gamma^0 \\ \frac{\partial \hat{p}}{\partial n} = 0 & \text{on} & \Gamma^1 \end{array} \qquad (11)$$

**Adjoint model**
By multiplying the tangent linear model by a function q and integrating by parts twice, we have:

$$\int_\Omega -div(K\nabla q)\hat{p} = \int_\Omega -div(k\nabla p)q$$

we consider

$$\begin{array}{rll} -div(K\nabla q) = 2(p - p_r) & \text{in} & \Omega \\ q = 0 & \text{on} & \Gamma^0 \\ \frac{\partial q}{\partial n} = 0 & \text{on} & \Gamma^1 \end{array} \qquad (12)$$

We have then

$$\int_\Omega 2(p - p_r).\hat{p} d\Omega = -\int_\Omega k\nabla p.\nabla q$$

as

$$(2(p - p_r), \hat{p}) = (\nabla J(K), k)$$

We have

$$\nabla J_K(k) = \int_\Omega -k\nabla p.\nabla q \qquad (13)$$

Finally

$$\frac{\partial J}{\partial K_i}(k) = -\int_{\Omega_i} k\nabla p.\nabla q, \ i = 1, \ldots, 8 \qquad (14)$$

In summary, to calculate the sensitivity, we proceed as follows for each set of parameters and for each parameter:

1. Solve the direct problem (3)
2. Solve the adjoint problem (12)
3. Calculate $S_i$ of parameter $K_i$ using equation (14)

As can be seen, the calculation of sensitivities of different parameters, requires several time the solving of problems (3) and (12). This requires enormous computing capacity if the number of parameters and the number of data sets are very high. To solve this

problem of resources, we ran the program on the Moroccan grid (magrid).

### 5.5 Grid Computing

#### 5.5.1 Definition

According to FOSTER [14], a grid computing is a system that:

- coordinates the resources which are not under the control of a central system,
- uses standards, opened and generic protocols and interfaces,
- provides multiple high quality services.

The ingenuity of the concept of the grid lies in its ability to virtualize resources. With this virtualization, we see the overall system as a super virtual machine. Grid computing allows you to have great computer's resources for the realization of heavier calculation.

#### 5.5.2 Grid computing operation:

To use a grid, you must have a certificate that allows you to attach a virtual organization which is a dynamic group of entities that choose to share resources and to define the conditions and roles of sharing them. Then we must create a proxy to submit the application called "job" written in a language called "job description language (JDL)". Very briefly, the elements involved in the care of a job submitted to a grid.
The Workload Management System (WMS) is consisted of the following elements:

- User Interface (UI): interface through which users access to the grid
- Computing Element (CE): represent the access point unified to resources of calculation, of the worker nodes which will be used by the grid for the execution of the jobs. The Computing Element is responsible for the management of jobs assigned to it. the Computing Element maintains a list of jobs to submit (batch queue). Storage Element (SE) : manages the storage of information.
- Information System (IS): set of resource information indicating the characteristics and condition of the Computing Element (CE) and Storage Elements (SE)
- Resource Broker (RB) or Workload Manager (WM): matches the needs of users with the resources available on the grid.
- Worker Nodes (WN): group of machines on which jobs will be executed. It is also on the worker nodes that are stored the data from the Storage Element. This is usually a cluster of several computers.

For more details on the use of grid computing, see [15],[16].

### 5.6 Distributed programing and execution of our application on magrid

To launch our application on magrid, we started by installing Freefem++ on the grid with the help of researchers from the National Center for Scientific Research and Technology of morocco. Then we prepared our job for submission. To exploit the distributed architecture of the grid, we used a parametric job. Run a Parametric job is to run N jobs differing only by the value of a parameter which can indicate PARAM values. As we have two simulations, only the input file parameter changes. So PARAM shows the data file to use(see sensitivity.jdl of appendicies). To see [17] for magrid using.

### 5.7 Numerical results

To realize the sensitivity analysis, we randomly generated 10000 parameter set Ki, i = 1, ..., 8 following a uniform distribution on the interval [ai, bi] (range of parameter variations Ki, i = {1, ..., 8}). The reference values are those used in the simulation of groundwater level in 2000. We repeated two times the simulations. The estimated execution time on a personal computer hp intel (R) Core (TM) i5 CPU 540@2.53GHz M 2.53 GHz and 2 GB of RAM was 49 hours. On the grid, this time was reduced to 19 hours. The results are given in Figures 14 and 15.

By observing the sensitivity curves, we can see that four elements strongly influence the level of the water. These are the parameters 3 to 6 ie fractures 3 to 6. And particularly the fractures 3 and 6 (see Figure 1) whose the sensitivities are too high.

## 6 Conclusion

This work answers questions posed in the thesis of Mr Koussoubé[1]; it treats the numerical simulation of a flow in fractured porous media whose hydraulic conductivities are lower than those of the rock. We refined the mesh at the fractures that were treated as porous media. Our simulations have helped provide some answers on a real problem posed in the thesis of Mr Koussoubé[1]. A sensitivity analysis allowed identifying the parameters that influence a lot the solution. Collaboration with researchers of the National Center for Scientific Research and Technology of Morocco

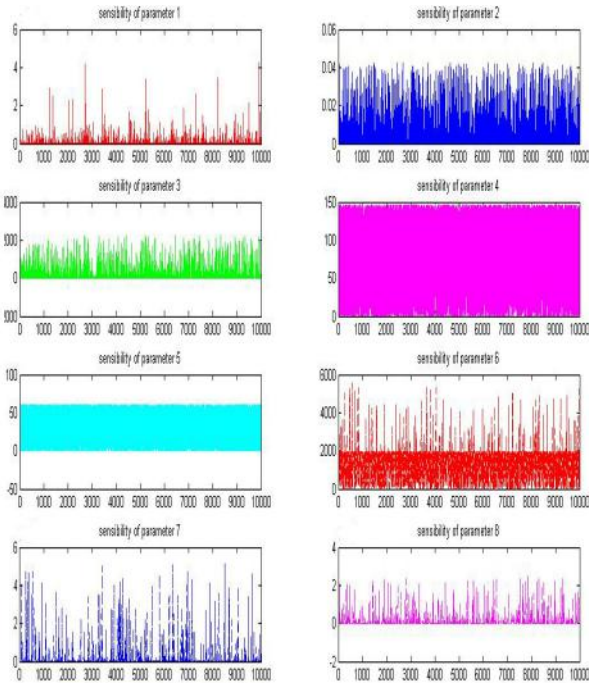



Figure 14: Analysis 2

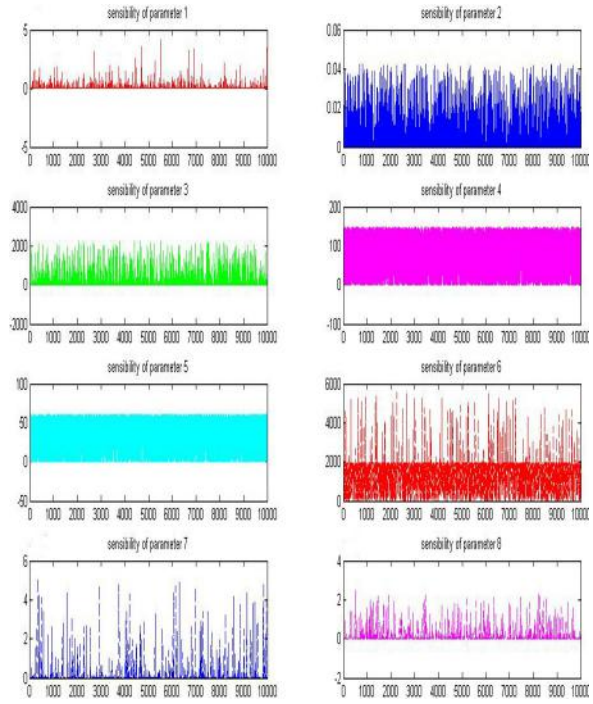



Figure 15: Analysis 3

and of the Meraka Institute of South Africain in the framework of project ”brain gain” of the UNESCO-HP has allowed us to install the software for numerical Freefem++ on the Moroccan grid (margrid) and the South African grid (sagrid). This gave us more substantial computing resources for our calculations.

We are planning thereafter to parallelize our computer code to perform simulations on the transfer of pollutants in the water. This will undoubtedly help install bore wells in order drinking supply water of the population.

## Appendix

**freefem.sh**

```
#!/bin/bash
source $VO_MAGRID_SW_DIR/setenv_freefempp_v3.14.sh
FreeFem++ -nw $1
tar -czf output$2.tar.gz s1.dat s2.dat
s3.dat s4.dat s5.dat s6.dat s7.dat s8.dat
```

**sensitivity.jdl**

```
Executable="freefem.sh";
JobType = "Parametric";
Parameters = 3;
ParameterStart = 1;
ParameterStep = 1;
Arguments="sensitivity_PARAM_.edp _PARAM_";
InputSandbox={"freefem.sh","sensitivity_PARAM_.edp",
"parameter_PARAM_.dat"}; StdOutput="sensitivity_PARAM_.out";
StdError="sensitivity_PARAM_.err";
OutputSandbox={"sensitivity_PARAM_.out",
"sensitivity_PARAM_.err","output_PARAM_.tar.gz"};
MyProxyServer="myproxy.ct.infn.it";
```

## Acknowledgements

The first author thanks the Agence Universitaire de la Francophonie (AUF) for its financial support, Unescohp through the project ”brain gain” that allowed training on the use of grid computing. He expressed his gratitude to Mr Youssouf Koussoubé of Laboratory of Hydrogeology at the University of Ouagadougou for the clarification. Finally the authors gratefully acknowledge the administrators of magrid of National Center for Scientific Research and Technology of Morocco and Mr Bruce Becker of the Meraka Institute in South Africa for its support in the installation of FreeFem++ on magrid and on Sagrid.

## References

[1] Y. Koussoubé, Hydrogéologie des séries sédimentaires de dépression piézométrique de Gondon(Bassin du Sourou)-E

Faso,Thèse de doctorat de l'université Pierre et Marie Curie, Spécialité: Hydrogéologie, juillet 2010.

[2] F. Hetch, O. Pironneau, FreeFem++, http://www.freefem.org.

[3] E. Ledoux, Modèles mathématiques en hydrogéologie, Centre d'Informatique Géologique ´Ecole Nationale Supérieure des Mines de Paris, 2003.

[4] C. Alboin, J. Jaffré, J. E. Roberts, Domain decomposition for some transmission problems in flow in porous media , In Numerical treatment of multiphase flows in porous media, (Beijing, 1999), volume 552 of Lecture Notes in Phys., pages 2234. Springer, Berlin, 2000.

[5] T. Arbogast, The double porosity model for single phase flow in naturally fractured reservoirs , In M.F. Wheeler., editor, Numerical simulation in oil recovery, volume 11 of the IMA volumes in Mathematics and its Applications, pages 23-45. Springers Verlag 1988.

[6] H. MUSTAPHA Simulation numérique de l'écoulement dans des milieux fracturés tridimensionnels, Thèse de l'Universitéde Rennes 1, 2005.

[7] G. de Marsily, Cours d'hydrogéologie ,Université Paris VI, Septembre 2004.

[8] C. Alboin, Deux outils mathématiques pour modéliser l'écoulement et le transport de polluants dans un milieux poreux saturé, Thèse de l'UniversitéParis IX Dauphine, 2000.

[9] P. G. Ciarlet , Introduction à l'Analyse Numérique Matricielle et à l'Optimisation, Masson, Paris, 1982.

[10] P. G. Ciarlet , The Finite Element Method for Elliptic Problems, North-Holland,1980.

[11] V. Martin, Simulations multidomaines des écoulements en milieu poreux, Thèse de l'UniversitéParis IX Dauphine,2004.

[12] C. Lauvernet , Assimilation variationnelle d'observations de télédétection dans les modèles de fonctionnement de la végétation: utilisation du modèle adjoint et prise en compte de contraintes spatiales, Doctorat de l'UniversitéJoseph Fourier-Grenoble 1, Avril 2005.

[13] H. E. Ngodock, Assimilation de données et Analyse de sensibilité: Une application à la circulation océanique., Doctorat de l'UniversitéJoseph Fourier-Grenoble 1, Avril 2005..

[14] I. Foster, What is the Grid ? A three point check-list, Chicago, Argonne National Laboratory and University of Chicago, 2002 .

[15] GILDA , gLite users tutorial, https ://grid.ct.infn.it/twiki/bin/view/GILDA/ UserTutorials.

[16] B. Jacob, Brown, K. Fukui, N. Trivedi, Introduction to Grid computing, Redbooks, 2005.

[17] CNRST Rabat, Utilisation Magrid , http://wiki.marwan.ma.

**First Author** is a PhD. candidate in applied mathematics at the University of Ouagadougou(Burkina Faso) since 2009, current research is about mathematic modelling and numerical Simulation of flow in porous medium.
**Second Author** received his Master of Science and his Ph.D. degrees from the University of Nancy I France respectively in 1986 and 1989. In 2006, he received the HDR in Applied Mathematics from the University of Cadi Ayyad, Morocco. He is currently Professor of modeling and scientific computing at the Faculty of Sciences and Technology of Marrakech. His research is geared towards non-linear mathematical models and their analysis and digital processing applications.
**Third author** is a professor at the University of Ouagadougou since 1984 and is University Professor of CAMES since 2004. He directs the Laboratory of Numerical Analysis, Computer Science and Biomathematics. His research focuses on numerical simulation and mathematical modeling.

# Implementation of Location based Services in Android using GPS and Web Services

**Manav Singhal[1], Anupam Shukla[2]**

**[1]ABV-Indian Institute of Information Technology and Management
Gwalior, India**

**[2]ABV-Indian Institute of Information Technology and Management
Gwalior, India**

**Abstract**

Location based Services offer many advantages to the mobile users to retrieve the information about their current location and process that data to get more useful information near to their location. With the help of A-GPS in phones and through Web Services using GPRS, Location based Services can be implemented on Android based smart phones to provide these value-added services: advising clients of current traffic conditions, providing routing information, helping them find nearby hotels.

In this paper, we propose the implementation of Location based services through Google Web Services and Walk Score Transit APIs on Android Phones to give multiple services to the user based on their Location.

**Keywords** - *Android Mobile Operating System, Location Based Services, Web Services, A-GPS*

## 1. Introduction

The idea of using the mobile handsets and phones is to deliver the valuable services except the basic communication that had been started in the early 1990s when Internet was added in Voice Telephony.

Location-based services or LBS [1] refer to 'a set of applications that exploit the knowledge of the geographical position of a mobile device in order to provide services based on that information.'

Location-based services (LBS) provide the mobile clients personalized services according to their current location. They also open a new area for developers, cellular service network operators, and service providers to develop and provide value-added services: advising clients of current traffic conditions, providing routing information, helping the users to find nearby shopping malls.

Location-based services offer many merits to the mobile clients. For the mobile user, the examples of location-based services [2] are:

- To determine the nearest business or service, such as an Bank or Hotels
- Receiving alerts, such as notification of Sale in Shopping Mall or news of Traffic Jam nearby.
- Friend finder or receiving the location of the stolen phone.

Location based Services can be classified in 3 categories [1]-

a) Public Safety / Emergency Services

The location of the client can be determined by the mobile carrier hence it finds great use during Emergency since it can be used during the emergency/health hazard to locate the mobile clients.

b) Consumer Services

Now days, smart phones like (Android, Blackberry and iPhone) provide a set of location based applications and services which helps the users to access the multiple services based on the user location.

- *Maps Navigation*- The users can use the Google Maps to get to the particular location or to trace the route between any two locations.

- *Marketing /Advertising*- Many corporate companies advertise their items based on the location of the clients.
  For Example – Sale in Shopping Mall near to your location.

- *Location based Reminders*- The phones can be used to set as the reminder based on the location.
  For e.g. - Setting the Location based Alarm while traveling in the train

- *Preferred Location Search*- The user can also initiate the search of any nearby ATM or Restaurant within 5/10/15 kms range from his current present location.

There are two methodologies to implement LBS [3]-

- To process location data in a server and to forward the generated response to the clients.

- To find location data for a mobile device-based application that can use it directly.

To discover the position of the mobile, LBS must use positioning methods in real time. The accuracy of the methodology depends on the approach used. Locations can be represented in spatial terms or as text descriptions.

A *spatial location [2]* can be represented in the used latitude-longitude-altitude coordinate system. Latitude is defined as 0-90 degrees north or south of the equator and longitude as 0-180 degrees east or west of the prime meridian, that passes through the Greenwich, England. Altitude is represented in meters above sea level.

A *text description* is usually defined as a street location, including city, pin code.

The location of the device can be retrieved by-

i) Mobile Phone Service Provider Network-

The current cell ID is used to locate the Base Transceiver Station (BTS) that the mobile phone is interacting with and the location of that BTS. It is the most basic and cheapest method for this purpose as it uses the location of the radio base station that the cell phone is connected to.

A GSM cell may be anywhere from 2 to 20 kilometers in diameter. Other approaches used along with cell ID can achieve location granularity within 150 meters. The granularity of location information is poor due to Wide Cell Range. The advantage is that no additional cost is attached to the handset or to the network to enable this service.

ii) Satellites

The Global Positioning System (GPS) uses a constellation of 24 satellites orbiting the earth. GPS finds the user position by calculating differences in the times the signals, from different satellites, take to reach the receiver. GPS signals are decoded, so the smart phone must have in-built GPS receiver.

Assisted-GPS (A-GPS) is the new technology for smart phones that integrates the mobile network with the GPS to give a better accuracy of 5 to 10 meters. This fixes the position within seconds, has better coverage and can, in some cases, be used inside the buildings, consumes less battery power and requires fewer satellites.

The granularity of location information is most accurate (Latitudes and Longitudes).The disadvantage is cost of A-GPS enabled handsets for the user.

## 2. Background

In the last few years, the smart phones (Android, Black berry and iPhone) have taken over the market of Nokia based Symbian Phones in India. And these smart phones come equipped with A-GPS functionality which provides the spatial coordinates of the user location.

Android's Network Location Provider determines user location using cell tower and Wi-Fi signals, providing location information in a way that works indoor and outdoor, responds faster, and uses less battery power.

Assisted GPS [6], also known as A-GPS or AGPS, improves the performance of standard GPS in devices connected to the wireless network. A-GPS enhances the location granularity of cell phones (and other connected devices) in two ways:

- By helping in finding a faster "time to first fix" (TTFF). A-GPS acquires and stores information about the location of satellites via the cellular network hence the information does not need to be downloaded via satellite.

- By helping position mobile device when GPS signals are not strong or not present. GPS satellite signals may be impeded by tall towers, and they do not penetrate building interiors well. A-GPS uses proximity to cellular towers to calculate location when GPS signals are unavailable.

It addresses signal and wireless network problems by using assistance from other services. Such a technology in our smart phones can assist in various ways like tracking current location, receiving turn-by-turn direction instructions, route tracking, etc.

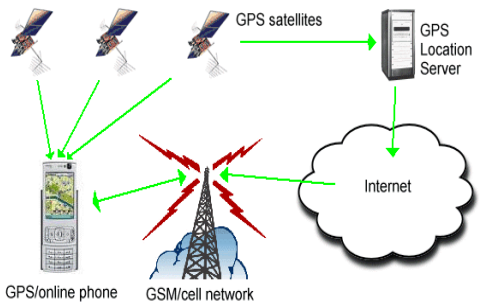



Figure 1 Architecture of A-GPS System

Mostly suited for mobile devices, A-GPS takes assistance from GPRS and at times, the service provider network information, to pin-point the current location accurately. Moreover the amount of CPU and programming required for a GPS phone is reduced by diverting most of the work to the assistance server instead.

A typical A-GPS enabled cell phone uses GPRS or other such Internet based data connection to build a contact with the assistance server for A-GPS. As this technique does not take into account the cell phone service provider network completely, we only pay for the GPRS usage charges and nothing else. The only down-side to this technology is that an A-GPS server cannot utilize any of the three standby satellites available for GPS connections.

AGPS minimizes the amount of memory and hardware that must be integrated into mobile devices in order to provide GPS-quality device locating ability as required by mobile devices. This keeps the mobile device simple and allows longer battery time.

GPS is real-time solution provider whereas AGPS is not. The network usage is required every time we move out of the service area. It is useful only for locating a particular place in small area. There is no privacy in GPS and A-GPS since the Assistance server knows the location of the device.
There needs to be communication over the wireless for processing of GPS information so this could be expensive.

## 3. Implementation and Methodology

Location-based service is another key functionality that gets used in smart phone applications. It is often combined with maps to give a good experience to the user about their location.

Android support LBS Application Programming Interfaces (APIs) [7]. Location service allows finding out the device current location. The application can request for periodic update of the device location information. The application can also register a intent receiver for proximity alerts like when the device is entering and existing from an area of given longitude, latitude and radius.

### 3.1 Android Location API

These are the different classes present under Location API package to retrieve the Location information of the user. [7]

- *LocationManager*- The class provides access to the location service. It also provides facility to get the best Location Provider as per the criteria.
- *LocationProvider*- It's an abstract super class for location providers. A location provider provides periodic reports on the geographical location of the device.
- *LocationListener*- This class provides callback methods which are called when location gets changed. The listener object has to be registered with the location manager.
- *Criteria*- The class provides the application to choose suitable Location Provider by providing access to set of required properties of the LocationProvider.

Android also provide an API to access the google maps. So with the help of the google maps and the location APIs the application can show required places to the user on the map.

### 3.2 Google Places API

On 10 May, 2011, at the Google I/O developer Conference in San Francisco, Google announced the opening up and general availability of the Google Places API.

The Google Places API [8] is a service that returns data about Places — defined within this Web Service as, spatial locations, or preferred points of interest — using HTTP requests. Place response specifies locations as latitude/longitude coordinates.

The four types of requests are available with the Google Places API-

There are 4 fundamental Place services available:

- *Place Searches* - It returns an array of nearby Places based on a location defined by the user.
- *Place Details* - It returns more specific data about a user defined Place.
- *Place Check-ins* - It allows the request that a person has checked in to a Place. Check-ins is used to gauge a Place's popularity; frequent check-ins will boost a

Place's priority in application's Place Search responses.
- *Place Reports* - It allows the users to add new locations to the Place service, and to delete Places that the application has added to the database.

The Google Places API [8] has the following limitations on the query processing:

- Users are allowed only 1000 requests per 24 hour period which are having the API Key.
- Clients who have also validated their identity through the APIs console are allowed 100 000 requests for 24 hours period. A credit card is required for authentication, for enabling billing.

### 3.2.1 Place Searches

A Place Search request is an HTTP URL defined in the following way [8]:

*https://maps.googleapis.com/maps/api/place/search/output?arguments*

Where output may be either of the following values

- json shows the response in JavaScript Object Notation (JSON)
- xml shows output as XML.

Table 1: Place Search API Arguments

| Arguments | Description |
|---|---|
| Location (required) | The latitude/longitude about which place information is to be found out. This must be defined as latitude, longitude. |
| *Radius (required)* | Distance (in meters) about which to show Place results. |
| *types (optional)* | Limit the results to places matching at least one of the pre defined types. Types must be separated with a pipe notation (type1\|type2\|etc). |
| *Language (optional)* | The language code, showing in which language the results must be shown, if possible. |
| *name (optional)* | A term to be mapped against the names of Places. Results will be limited to those having the name. |
| *sensor (required)* | Indicates whether or not the place request is from the device having a location sensor (e.g. a GPS) to find the location sent in this request. This value is either true or false. |
| *key (required)* | Application's API key. The key determines your application's identity so that places added from the application are made available immediately. |

### 3.2.2 Place Details

A Place Details [8] request returns more detailed information about the user defined place such as its address, contact number, user rating, etc.

Once we have a Reference Number of Particular Place from Place Search Request, we can initiate the search about that place details.

A Place Details request is an HTTP URL of the following form:

*https://maps.googleapis.com/maps/api/place/details/output?arguements*

- json (recommended) shows the output in JSON
- xml gives output as XML.

Table 2: Place Detail Web Service Arguments

| Arguments | Description |
|---|---|
| *reference (required)* | A identifier that uniquely defines a place, given from a Place search request. |
| *language (optional)* | The language code, showing in which language the results should be returned. |
| *sensor (required* | Defines whether or not the Place Details request is from the device having a location sensor (e.g. a GPS). This value is either true or false. |
| *key (required)* | The application's API key. This key identifies the application for purposes of quota management. |

## 3.3 Public Transportation API

The Public Transit API [10] from Walk Score gives the Transit Score for any location listed with in its database and provides convenient access to nearby public transit stops.

We can use the Public Transit Services to:

- To Add Transit Score to the application.
- To View public transit stoppages on a map
- To view the details about nearby transportation routes.

The Public Transit API has information database from over 200 public transit agencies in the world.

### 3.3.1 Stop Search API

The stop search API [10] call gives the data about the stoppages of the public transport near a given location. This call returns 16 stops that service unique routes near

the user defined location. We can say, each stop must contain a unique route.

For e.g., if there are 3 bus stops near a place, let's say New York that only served the Route 23 bus, only the first of those stops would be returned by the stop search API call.

Table 3- Stop Search API Request Parameters

| Parameter | Required | Description |
|---|---|---|
| *lat* | Yes | The latitude to search near. |
| *lon* | Yes | The longitude to search near. |
| *wsapikey* | Yes | Your Walk Score API Key. |

The response is returned in JSON or XML format with the following keys –

Table 4- Stop Search API Response Keys

| Key | Description |
|---|---|
| id | The stop's id. This can be used to query for further stop details. |
| lat | The stop's latitude. |
| lon | The stop's longitude. |
| name | The stop's name; often, this is the street intersection. |
| distance | The stop's distance to the search location point, in miles. |
| summary_text | A text summary of the stop, its distance to the search location. |
| summary_html | An HTML summary of the stop, suitable for adding directly to web page. |
| route_summary | JSON array of routes corresponding to this stop. |

E.g. - To get 16 bus stops near Pike Place Market that service distinct routes, make the following request.

*http://transit.walkscore.com/transit/search/stops/?lat=47.6101359&lon=122.3420567&wsapikey=key*

These were the APIs which can be used in different ways to provide the services based on the location of the user.

## 4. System Testing

We developed the mobile application on Android covering all the mentioned APIs and the application was tested using Samsung Galaxy S handset (which is A-GPS enabled handset).

Android Version – 2.1 (Eclair)

Android Permissions-
android.permission.INTERNET
android.permission.ACCESS_FINE_LOCATION
android.permission.ACCESS_COARSE_LOCATION

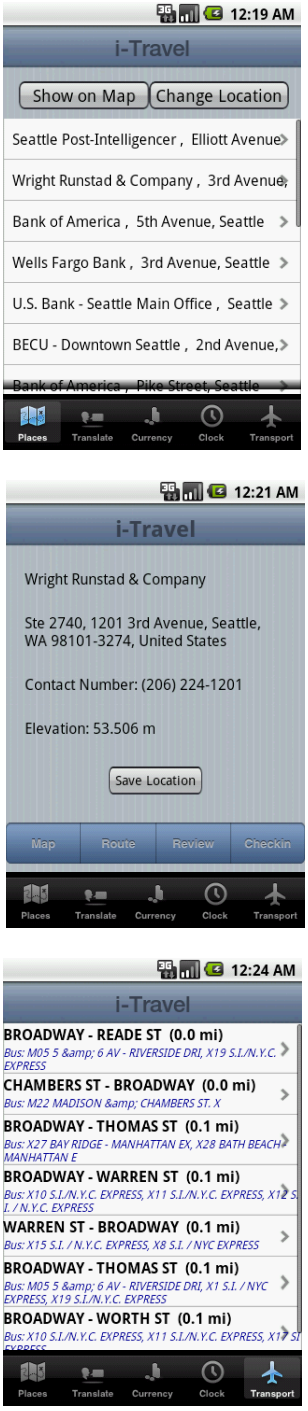



Figure 2- Screen showing a) Place Search Results b) Place Details using Google Places API c) Public Transit using Walk Score API

## 5. Conclusion

There are various constraints to implement Location Based Services. The different kinds of constraints include [1]:

- Technology Constraints

For LBS to be operational on a large scale, mapping under the geographical information system (GIS) needs to be more comprehensive than it is today. This raises significant challenges in for improving the breadth and the depth of the existing coverage of GIS. The most important factor in enabling the growth of LBS is wide availability of cheap GPS enabled handsets. GPS enabled handsets are being manufactured now days. The issue of cost remains to be tackled, since these phones are still all high-end units.

- Infrastructure Constraints

One of the main problems is the lack of spread of the wireless network into the countryside. In developing country like India, the wireless technology is in very nascent stage. In metro cities and areas, the problem of network congestion is also an important issue. The percentage of service operators not meeting the congestion rate benchmarks has risen substantially.

- Market failure

One of the main constraints to the provision of value added services, in general, and LBS in particular, is the market structure of the mobile industry and the failure to unleash the forces of competition. A key essential need for LBS provision needs cross-network connections to be seamless, and the current practices go against a cooperative attitude for LBS provision.

## References

[1] Location Based Services on Mobile in India
For IAMAI - Version: 14 April 2008
http://www.iamai.in/Upload/policy/LBS_Draft_Indicus.pdf

[2] J2ME and Location based Services
By Qusay H. Mahmoud - March 2004
http://developers.sun.com/mobility/apis/articles/location

[3] Location Based Services
By Valerie Bennett
http://www.ibm.com/developerworks/ibm/library/i-lbs

[4] Android Wireless Application Development
By Shane Condor and Lauren Darcy

[5] GPS Signal Acquisition and Tracking – An Approach towards Development of Software based GPS Receiver
By Dinesh Manandhar, Yongcheol Suh, Ryosuke Shibasaki

[6] WebServices.org Home Page
http://www.webservices.org

[7] Location Manager APIs– Android Developer
http://developer.android.com/reference/android/location/LocationManager.html

[8] Google Places API
http://code.google.com/apis/maps/documentation/places/

[9] Google Maps API
http://code.google.com/apis/maps/documentation/imageapis/index.html

[10] Walk Score Transit API
http://www.walkscore.com/professional/public-transit-api.php

[11] Google Geo Coding APIs
http://code.google.com/apis/maps/documentation/geocoding

[12] Location Management for Mobile Devices
Erik Wilde (School of Information, UC Berkeley) - February 2008
http://dret.net/netdret/docs/wilde-irep08-016-mobile-location.pdf

[13] Query Processing in Mobile Environments: a Survey and open Problems
N. Marsit, A. Hameurlain, Z. Mammeri, F. Morvan

[14] Location the Portal on positioning and navigation
www.location.net.in

[15] LBS Zone
www.lbszone.com

[16] Android Wireless Application Development
By Shane Condor and Lauren Darcy

Manav Singhal is pursuing Integrated Post Graduation in Information and Communication Technology from ABV-Indian Institute of Information Technology and Management, Gwalior. He is National Talent Search Exam (NTSE) scholar and his areas of interests include Mobile computing, Cloud Computing and had won many National awards in Android Application Development. He won Best Application Awards in Social and Business Category in Indian Android Developer Contest 2011, India.

Anupam Shukla is the professor in the ICT department of ABV-Indian Institute of Information Technology and Management Gwalior. He received his Ph.D. in Electronics & Telecommunication Engineering in 2002 from NIT Raipur. He has published around 70 papers in various national and international journals/conferences.

# MC-CDMA Scheme in Wi-Fi Environment

**N. Larbi[1], F. Debbat[2] and A. Boudghen Stambouli[3]**

**[1]Space Techniques Center - CTS**
**01 Avenue de la Palestine, PB 13, Arzew 31200, Algeria**

**[2] Computer Science Department - Faculty of Sciences and technology**
**University of Mascara 29000- Algeria**

**[3] Electronics Department - Faculty of Electrical and Electronic-University of Oran - USTO**
**Oran 31000–Algeria**

**Abstract**

The combination of OFDM and Code Division Multiple Access (CDMA) is seen as an attractive and practical solution to enhance the throughput and/or robustness for future high-speed indoor WLANs. The multi-path Rayleigh channel represents a hostile environment for WLANs communication. So, we have proposed the MC-CDMA system to overcome the impact of this kind of wireless channel.

The focus of this article is to simulate a modified physical layer (PHY) based on the IEEE 802.11a combined with a stage of spread spectrum. This modified layer particularly concentrates on IEEE 802.11a standard. Basically, the proposed schemes can be split in different types depending on the CDMA code used. We investigated the modified physical layer performance on the basis of Bit Error Rate (BER) and Signal-to-Noise Ratio (SNR). The numerical results show that the MC-CDMA is a powerful multi-carrier multiple access technology.

***Key words:*** *WLAN, IEEE 802.11a, OFDM, MC-CDMA, Rayleigh channel.*

## 1. Introduction

Wireless Local Area Networks (WLAN) have evolved as the quantity and types of mobile devices have increased. The number of devices that support Wi-Fi continues to expand. The term Wi-Fi is an industry acronym meaning wireless fidelity for devices with support for the IEEE 802.11 wireless standard. The type of devices that include Wi-Fi support continues to expand from laptops to many other devices, such as cameras, phones, automobiles, and other consumer devices. The uses of Wi-Fi have expanded beyond just data usage, often including voice, video, and innovative contextual applications [1].

The standards of WLANs that currently gain the most momentum are IEEE 802.11a and IEEE 802.11b. The IEEE 802.11a standard is based on OFDM. The main reason that OFDM was selected as basis for this standards is its capability to deal with the strong multi-path propagation present in indoor propagation channels. In general, OFDM splits a wideband frequency-selective fading channel into a number of narrowband frequency-flat fading channels (i.e., subcarriers). Moreover, the ability to include a proper guard interval between subsequent OFDM symbols provides an effective mechanism to handle Inter Symbol Interference (ISI) caused by severe multi-path propagation [2, 3].

The IEEE 802.11b standard applied direct sequence spread spectrum (DSSS) or frequency hoping spread spectrum (FHSS) and provided 1 or 2 Mbps. In 1999, the IEEE 802.11b was available. It was designed to operate in an indoor environment. For the 1 and 2 Mbit/s modulation, the Barker sequence shall be used as code [3, 4].

The combination of OFDM and CDMA is seen as an attractive and practical solution to enhance the throughput and/or robustness for future high-speed indoor WLANs.

The focus of this article is to simulate a modified physical layer (PHY) based on the IEEE 802.11a combined with a stage of spread spectrum. This modified layer particularly concentrates on IEEE 802.11a standard because its physical layer design is considered to be more robust against harsh propagation conditions. According to the IEEE specification the WLAN should transmit the data in frames, but the preamble and the header are not implemented in our studies. Only the part which contains the data to be transmitted is implemented with Matlab software. Basically, the proposed schemes can be split in different types depending on the CDMA

code used. The numerical results show that the MC-CDMA is a powerful multi-carrier multiple access technology.
The rest of the paper is organized as follows: Section 2 gives a short overview of Wi-Fi technologies. Section 3 provides the background information for the MC-CDMA technique. In Section 4 a detailed presentation of the proposed scheme is given. Section 5 contains a discussion on simulation results. Section 6 summarizes this work with concluding remarks.

## 2. Overview of Wi-Fi technologies

In June 1997 the Institute of Electrical and Electronics Engineers (IEEE) defined an international interoperability standard, called IEEE 802.11 wireless LAN, also known as Wi-Fi. This standard specifies a number of Physical Layers (PHYs). Two of these PHY are based on radio communication and use the 2.4GHz band, license free ISM (Industrial, Scientific and Medical bands) and the others PHY uses infrared light. All three PHYs support a data rate of 1Mbps and optionally 2 Mbps.

In 1998, Lucent Technologies and Harris Semiconductor proposed a standard called Complementary Code Keying (CCK) to achieve 5.5Mbps and 11Mbps transmit rates. The IEEE adopted the CCK and released a new standard, named IEEE 802.b, in 1999.

Motivated by the demand for higher data rates and by the opening of new unlicensed spectrum in the 5GHz band for the use of a new category of equipment called Unlicensed National Information Infrastructure (UNII) devices, a new IEEE 802.11 working group started working on third generation of WLANs. In July 1998, this group selected OFDM as a transmission technique. In 2000, the standard was ratified and called IEEE 802.11a. It defines data rates between 6 and 54Mbps. To make sure that these data rates are also available in 2.4GHz band, mid 2003 IEEE standardisation group finalised a similar standard for this band named IEEE 802.11g. The growing success of IEEE 802.11b and IEEE 802.11a products together with the demand for even higher bit rates confirms the need for research to high data-rate extensions for WLANs.

## 3. MC-CDMA techniques

Multicarrier code-division multiple-access (MC-CDMA) represents a fusion of two radio access techniques, namely OFDM and the CDMA [5]. Such a combination has the benefits of both OFDM and CDMA [6]. In MC-CDMA symbols are modulated on many subcarriers to introduce frequency diversity instead of using only one carrier like CDMA. Thus MC-CDMA is robust against deep frequency selective fading compared to DS-CDMA [7]. Each user data is first spread using a given high rate spreading code in the frequency domain. A fraction of a symbol, corresponding to a chip of the spreading code, is transmitted through a different subcarrier. Hence each OFDM subcarrier has a data rate identical to the original input data rate. The transmitted signal of the i$^{th}$ data symbol of the j$^{th}$ user $s_i^j(t)$ is:

$$s_i^j(t) = \sum_{k=0}^{N-1} b_i^j c_k^j e^{2\pi(f_0 + kf_d)t} p(t - iT) \quad (1)$$

where: $N$ is the number of subcarriers, $b_i^j$ is the i$^{th}$ message symbol of the j$^{th}$ user, $c_k^j$ represents the k$^{th}$ chip, $k = 0,1,..., N-1$, of the spreading sequence of the j$^{th}$ user, $f_0$ is the lowest subcarrier frequency, $f_d$ is the subcarrier separation and $p(t)$ is a rectangular signaling pulse shifted in time given by:

$$p(t) = \begin{cases} 1 & 0 \le t \le T \\ 0 & elsewhere \end{cases} \quad (2)$$

If 1/T is used for $f_d$, the transmitted signal can be generated using the IFFT, as in the case of an OFDM system. There for, if the original symbol rate is high enough to become subject to frequency selective fading, the input data have to be serial-to-parallel converted into P parallel data sequences and each serial-to-parallel output is multiplied with the spreading code of length N. Then each sequence is modulated using Ns subcarriers. Thus, all Ns = P.N subcarriers are also modulated in baseband by the IFFT.

## 4. Proposed schemes description

### 4.1 MC-CDMA model 1

In our simulations, the channel coding is not used. The proposed system uses MC-CDMA in the physical layer. As shown in Figure 1, first the input data stream is modulated with BPSK modulation.
Then the modulated data symbol sequence is serial-to-parallel converted to a maximum of P=5 parallel sequences. Each of the parallel sequences is duplicated into spreading factor parallel copies and each of the duplicated symbols is multiplied by a chip from the spreading code. As spreading sequences the orthogonal Barker code of length N=11 is used.

After that an IFFT is performed and the pilot signal is used. Then a cyclic prefix is inserted after being converted to serial data stream to generate the MC-CDMA signal.
Finally, the signal is transmitted trough the channel. The channel consists of a multi-path fading with AWGN (10-Tap Rayleigh fading model is used as the channel); at the receiver the inverse operation is employed.

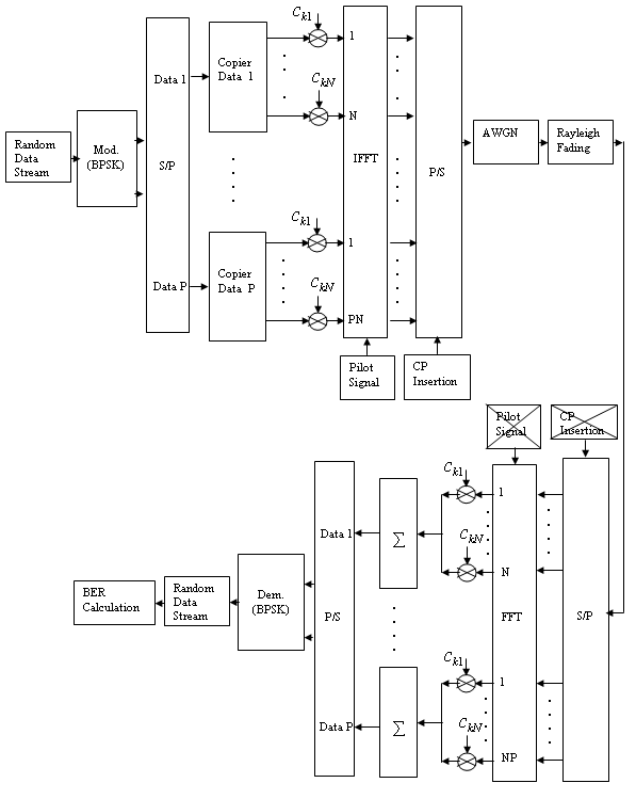



Fig.1 The first model simulation block diagram of MC-CDMA in Wi-Fi environment

## 4.2 MC-CDMA model 2

The 20 MHz channel is split into 52 subcarriers (48 subcarriers for data and 4 pilot subcarriers), like in the IEEE 802.11a OFDM physical layer.
Figure 2 illustrates the second MC-CDMA simulation model. The input data stream is first mapped into BPSK and then is multiplied by the spreading code of length N (the same symbol is transmitted in parallel through several subcarriers). Each chip of PN code modulates one subcarrier. The number of subcarriers in this scheme equals the length of spreading code.
All data corresponding to the total number of subcarriers are modulated in baseband by IFFT and converted back into serial data. Then a cyclic prefix is inserted between the symbols to combat the inter-symbol interference (ISI) and the inter-carrier interference (ICI) caused by multi-path fading.

Finally, the data are sent to the receiver over the same channels used before. At the receiver the inverse operation is employed.

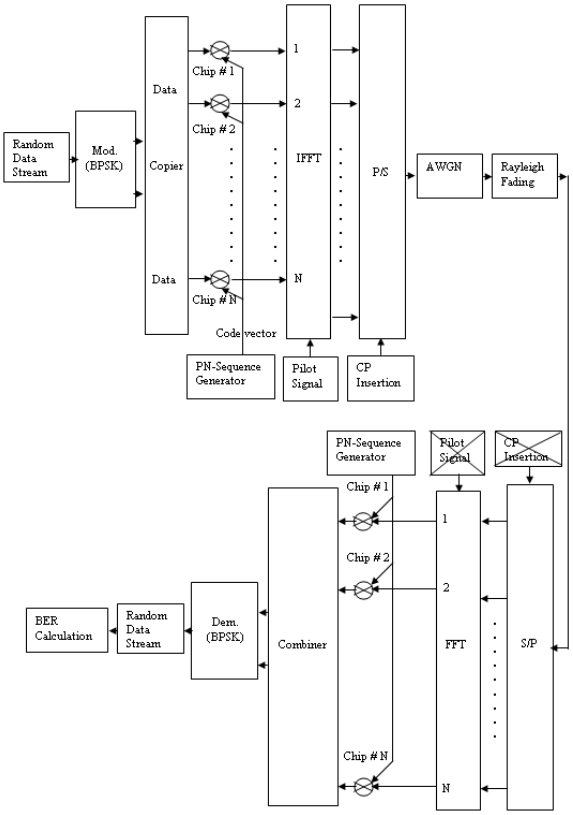



Fig.2 The simulation block diagram of MC-CDMA (2) in Wi-Fi environment

## 4.3 simulation parameters of the MC-CDMA models

Additional parameters of the PHY layer are presented in table 1.

Table 1. Simulation parameters of the MC-CDMA models

| ***Parameter*** | ***Value*** | | |
|---|---|---|---|
| | Model 1 | Model 2 | |
| | | Case 1 | Case 2 |
| Channel Bandwidth | 20 MHz | | |
| Subcarriers Spacing | 0.3125 MHz | | |
| Symbol Interval | 4μs = 3.2μs + 0.8μs | | |
| Guard Interval | 0.8μs | | |
| FFT size | 64 | 64 | 128 |
| Number of subcarriers | 55 | 52 | 104 |
| Spreading factor | 11 | 52 | 104 |
| Spreading code | Barker | PN | PN |
| P: parallel data sequences | P=5 | - | - |

## 5. Simulations, results and analysis

### 5.1 Simulations environment

The simulations are designed and implemented using Matlab. Performance is evaluated by transmitting randomly generated data stream over a channel. The stream is then received, demodulated and BER is calculated each time for every simulation. Multi-path appears in conditions where the transmitted signal experiences reflections, diffractions, and scattering. This is due to obstacles between transmitter and receiver. A channel in mobile communications can be simulated in many different ways. In our simulations, we have considered the two most commonly used channels: the AWGN model and a Rayleigh fading channel model. The channel model is based on the worst case scenario, assuming that no line-of-sight path is available between the transmitter and the receiver. On the other hand, a wideband fading channel can be modeled as a sum of several differently delayed, independent Rayleigh fading process. The corresponding channel impulse response is described as [7]:

$$h(t,\tau) = \sum_{p=1}^{P} a_p . R_p(t) . \delta(\tau - \tau_p) \qquad (3)$$

where: $a_p$ is the normalized amplitude, $R_p(t)$ is the Rayleigh fading process, $\tau_p$ is the delay of the p[th] path.

### 5.2 MC-CDMA simulation results

For the simulations a multi-path Rayleigh fading channel model is presented, the graphs of BER versus Es/No for the different models are shown in figure 3. The first model is based on the Barker code and the second model is based on PN code.

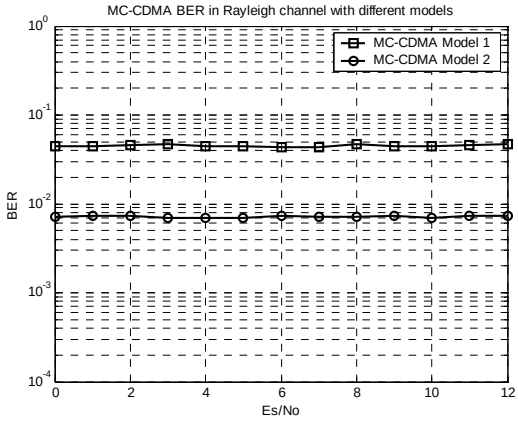



Fig.3 The BER vs. Es/No of MC-CDMA models in Rayleigh channel

The graphs indicate that the transmission quality is degraded in the Rayleigh channel. It is noted that the transmission quality is better for the second model.

Since the MC-CDMA model 2 presents a better performance then the first model, different subcarrier number are simulated. It is noted that the performances are influenced by the change of subcarrier number (FFT=64, FFT=128) as indicated in figure 4. The higher subcarriers number is better in BER.

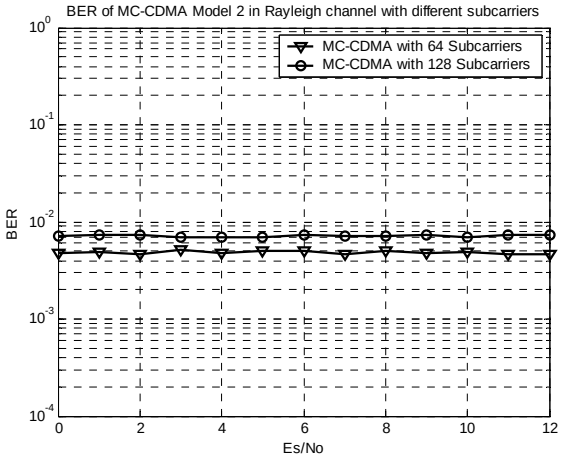



Fig.4 The BER vs. Es/No of the second MC-CDMA model in Rayleigh channel with 64 and 128 subcarriers

## 6. Conclusions

With the rapid development of wireless communications, two trends are gaining in popularity. One is the use of CDMA and the other is the use of OFDM. For future wireless applications, researchers have also started a new trend of combining these two techniques with the intention of obtaining advantages of both, and the proposed new technique is generally termed MC-CDMA. In this paper, we are proposing a modified version of the IEEE 802.11 system, based on MC-CDMA as PHY layer. In our simulations, we considered two different models. In order to verify the performances, we have evaluated them in multi-path Rayleigh fading channel which represent a hostile environment for WLANs communication. In each models, we measured the BER versus Es/No.

In the simulation results, it is interesting to note that:

- The transmission quality is degraded in the Rayleigh channel because of the multi path effects and the not use of the channel coding.
- The performances of the transmission are influenced by the choice of spreading code used (Barker, PN). The PN code shows a better transmission.
- The performances of the transmission are influenced by the change of subcarrier number (FFT=64, FFT=128). The higher subcarriers number is better in BER.

## References

[1] J. Smith, J. Woodhams, R. Marg, Controller-Based Wireless LAN Fundamentals, Cisco Systems, Inc, 2011.
[2] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: High-Speed Physical Layer in the 5 GHz Band, IEEE 802.11a, Sept. 1999.
[3] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE 802.11, 2007.
[4] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: High-Speed Physical Layer Extension in the 2.4 GHz Band, IEEE 802.11b, Sept. 1999.
[5] S. Hara, R. Prasad, Overview of multicarrier CDMA, IEEE Communications Magazine, vol. 36, December 1997, pp. 126–133.
[6] H. Schulze, C. Lüders, Theory and Applications of OFDM and CDMA Wideband Wireless Communications, John Wiley & Sons, Ltd, 2005.
[7] L. Hanzo, M. Munster, B. J. Choi and T. Keller, OFDM and MC-CDMA for broadband Multi-user communications, WLANs and broadcasting, John wiley & sons, LTd, 2003.

**Nacéra LARBI** is a researcher in Space Techniques Center (CTS), Algeria. As a member of power system research group and ALSAT-2 team (second Algerian satellite), her field of interests are focused on power electronics for micro satellites, power subsystems design and multiuser detection in the context of mobile communication using the combination of OFDM/CDMA. Nacéra LARBI received her bachelor degree in electronics in 1998, from Oran University of Science and Technology (USTO - Algeria), and her master degree in space technologies in 2002, from the Space Techniques Center, Algeria. She has 9 years experience on satellites power subsystem. Currently, she is pursuing her Ph.D. degree in the department of Electronics at the USTO University, Algeria.

**Fatima DEBBAT** received his bachelor degree, in 1996, from Oran University of Science and Technology USTO, and a master degree in space technologies, in 2002, from the Space Techniques Center (CTS) Algeria and Ph. D. degree in electronics, in 2007, from University of Telemcen, Algeria. She is currently a professor at Department of Computer Science, Mascara University, Algeria. Her current research interest covers Artificial intelligence applications, optimization, and wireless network.

# A new path algorithm for the weighted multi-graphs WMGPA: application to the Direct Topological Method

**Abderrahmane Euldji[1], Abderrahim Tienti[2] and Amine Boudghene Stambouli[3]**

**[1] National Institute of Telecommunications and ICT ( INT-TIC ), Oran, Algeria**

**[2] National Institute of Telecommunications and ICT ( INT-TIC ), Oran, Algeria**

**[3] Department of Electricity, University of Science and Technology of Oran USTO, Oran, Algeria**

**Abstract**

The aim of this paper is to present an algorithm which gives all the possible paths that start from a specific node to another of a weighted multi-graph. This algorithm is intended to be applied for the direct topological method.

***Keywords:*** Electrical circuits analysis, direct topological method, graph theory, path, weighted multi-graph, algorithm.

## 1. Introduction

The path notion is one of the most important graph traversal concepts. To find a path in a graph, two vertices shall be specified as a beginning and an ending vertex. The path is the chain that starts from the beginning vertex until it reaches the ending node, such that every vertex in this chain shall not be crossed twice. As a result, a path is a simple graph that does not contain any isolated node.

The path can be used to describe certain characteristics of a graph, or to solve other problems related to this latter. It can be used for decisional purposes, or for optimization needs. In general, many topics in the graph theory are based upon the path notion.

For this purpose, many efforts have been done in order to give an "optimized" algorithmic form that finds all the possible paths that starts from a specific node to another in a graph; the most well-known all paths extractor algorithms are the BFS (Breadth First Search), and the DFS (Depth First Search) and their derivative algorithms [1][2][3][4][5]. But since there are multiple classifications of graphs, then a path algorithm made for a specified class of graphs may or may not be used for the other classes of graph.

For electrical circuits -in which this paper is established for- the DFS at a first sight may appear as the appropriate procedure to find paths of the electrical circuit' associated graph, but since they are considered as weighted multi-graphs, then DFS becomes ineffective (incompetent). Hence, this paper proposes a path algorithm suited for the electrical circuits' analysis, and that can be used specifically for the direct topological method DTM; it is called: WMGPA.

Before everything, it is better suited to talk about the selected electrical circuits' analysis method in this paper, which is the direct topological method DTM.

## 2. Direct Topological Method DTM

The methods of analysis of the electrical circuits are divided into two main classes:

- The algebraic methods: based on the resolution of the circuits equations, either by conventional computations, or through the intermediary of the signal flow graph.
- The topological methods: they study the circuit's structure (topology) to deduct the expression of the circuit's functions.

The direct topological method (D.T.M) -as its name indicates- is a method of analysis of the permanent linear circuit networks, which permits to write directly the terms of the looked for circuit functions, under their definitive compact form, by visual examination of the studied circuit topology. While reducing the steps to get the result without complex computations, by simple application of the topological rules. Hence, it permits to solve a problem of circuit's analysis with minimum effort and economy of time in comparison with the other methods.

D.T.M is based on the graph theory definitions, since it is among the topological methods.

Any circuit function is rational; D.T.M intends to write separately the numerator and the denominator, while with the other analysis methods, only the circuit function is well determined. The denominator is called the topological determinant D; for the case of transfer functions, the numerator is called the topological transfer numerator N, both of these two parameters are called: the circuit's topological functions [6].

This method analyzes many types of electrical networks: passive or active circuits, circuits with dependent or independent sources, circuits containing one or many mutual inductances as well as ideal transformers. This method analyzes also circuits with distributed parameters as transmission lines. All of these circuits require -in one or many phases of their analysis process- the analysis of RLC circuits.

Before introducing how to analyze RLC circuits by the D.T.M, one should know some graph theory definitions.

## 3. Graph theory

Graph theory is the study of the properties of graph structures. It provides us with a language with which to talk about graphs. The key to solve many problems is identifying the fundamental graph-theoretic notion underlying the situation and then using classical algorithms to solve the resulting problem.

Graph theory is very wide domain, in constant evolution in either fundamental or applicative research; the applications are very numerous, which justifies an important research in algorithmic.

The graph theory offers a certain educational interest on the other hand. Indeed, the definitions are simple, and real problems of research can be posed as "mathematical games" whose playful formulation can cover big difficulties.

As the graph models very numerous situations, the proposed problems are of more "natural" form [7].

The multiplicity of the applications also explains the variety of the definitions, or of the variants of definitions. So an article of graph theory always must "fix the definitions"[8].

### 3.1 Relative definitions to topological graphs

***1)*** **node (vertex):** A node "v" is an extremity point or an intersection point between branches.

**2) Branch:** A branch "e" is a link between two nodes; there are two types of branches, oriented branches (it is called in this situation by arc) and non-oriented branches (it is called in this situation by edge) .

**3) Graph:** A graph "G" (v, e) is constituted from "v", a nonempty set of nodes, and from "e", a set of branches, where each branch from "e" is a connection between two nodes from "v"; nodes number is denoted |v|, and the branches number is |e| [3].

There exist many classifications for graphs; such as:

**a) Embedded vs. Topological graph:** A graph is embedded if the vertices and branches have been assigned geometric positions. Thus any drawing of a graph is an embedding, which may or may not have algorithmic significance. Occasionally, the structure of a graph is completely defined by the geometry of its embedding.

**b) Simple graph vs. Multi-graph:** A simple graph is a graph having no loops or multiple branches. In this case, each branch in E(G) can be specified by its endpoints u; v in V(G).

In contrast, a multi-graph is a graph that it may contain loops and at least one multiple branches (*two adjacent vertices are connected via multiple branches*)

**c) Directed graph (digraph / oriented graph):** a directed graph (digraph) is a graph where at least one branch in it is oriented (arcs).

**d) Weighted graph:** In weighted graphs, each branch (or vertex) of G is assigned a numerical (or symbolical) value, or weight.

**e) Labeled vs. Unlabeled graph:** In labeled graphs, each vertex is assigned a unique name or identifier to distinguish it from all other vertices [7].

**4) Node's degree of a non-oriented graph:** it is the number of non-looped branches connected to this node.

### 3.2 Links and attributes

**1) Loop:** There is a loop when the branch extremities correspond to the same node.
**2) Cycle:** A cycle is a closed path. A loop is a cycle with a single node and branch.

**3) Supplement:** supplement of a branch set [$e_1$, $e_2$,…, $e_n$] in a graph "G" is the resulted graph from the elimination

of these branches, each one is followed by the coincidence of correspondent extremities.

**4) Connected graph:** A graph "G" is called connected, if it is possible to find at least a path joining two arbitrary nodes from it. A non-connected graph is a degenerated one [6].

**5) Transfer Graph:**

**a) Transfer cycle:** There are two distinct branches from a graph; "g" is called: the transmitter, and "h" is called the receptor; these two branches form a couple K = (g , h) called the transfer couple. Each cycle from a graph containing at the same time "g" and "h" is a transfer cycle joining "g" and "h", noted (g → h). A graph containing transfer couples is a transfer graph.

**b) Sign of the cycle:** The sign of the transfer cycle joining g to h is +1 or -1 depending on whether g and h are in the same direction or in an opposite direction in the cycle.

## 4. Analysis of RLC circuits by the D.T.M

The analysis process is subdivided into many steps which depend on the specified network function: if it is a transfer function, than a calculus of a topological transfer numerator and a topological determinant is required. If it is a driving point function (input function) than two topological determinants are needed.

Both of the circuit's topological functions are calculated by following a set of steps, these steps vary according to the complexity of their associated graphs, if the graph is complex, then further rules and theorems are executed.

These rules and theorems are specified to extract either the topological determinant or the topological transfer numerator, for either of these two situations, some of these rules and theorems are elementary (like rule n°1 and theorem n°1), some are executed for special topologies, others are general.

In order to identify the procedure to obtain the topological functions, the class of these graphs shall be identified as well. The classification is important. The class of the graph implies the procedure to process it (this includes the algorithm(s) to apply). The two previously mentioned graphs to be extracted are both labeled and weighted topological multi-graphs. In the case of the transfer numerator's graph, it is considered as a directed graph, in contrast with the topological determinant's graph (it is a non-directed graph).

### 4.1 Topological Determinant D

To calculate the topological determinant of a RLC circuit, we consider the graph of its non-excited circuit (hence it is non-directed graph). The excitation sources in the circuit have to be replaced by short-circuits for voltage generators, and by opened circuits for current generators [6].

Afterwards, the circuit's topology is processed; it includes also the verification of the circuit's complexity.

If the graph is simple, than a table of elementary topological determinants (TED) is consulted, if this table does not contain this graph, or the graph is complex, than further rules and theorems are needed.

In this paper, we introduce only two rules [9]:

**1) Rule n° 1:** An inductance "L" must be considered as impedance "LS', and a capacitance "C" must be considered as admittance "CS". A resistance could be considered as impedance "R", or admittance "G".

**2) Rule n° 3:** The determinant of a degenerated circuit is null.

A circuit is degenerated when it is not connected, or when it contains a cycle which all its branches have a null impedance. This cycle corresponds to a degenerated node of the graph of the circuit.

### 4.2 Transfer Numerator

To get the transfer function, we must have besides the determinant the transfer numerator. Contrary to the determinant that only depends on the non-excited circuit, the transfer numerator depends not only on the placement of the excitation source, but also to the one of the output response. Therefore, we have then to consider the circuit's graph with the excitation source and the output response. This graph differs from the non-excited circuit's graph by the existence of a transmitter branch, which represents the excitation source, and a receiver branch, which represents the output response, which both of them are oriented branches. Therefore, this is the transfer graph of a single transfer couple; it can contain several transfer cycles joining the source excitation to the output response.
In this paper, we introduce only one theorem.

**Theorem n°3:** The transfer numerator of a RLC circuit is the algebraic sum of the circuit's transfer cycles values [9].

The value of the transfer cycle, as a definition, equals the multiplication of:

- The cycle's sign.
- Admittances being part of the cycle.
- The cycle's supplement determinant of the circuit.

## 5. Specification of the data structures

As already said, the graphs classification have a major impact on the choice of the procedure/algorithm to be applied. This also have an important role for the choice of the most appropriate data structures to be selected.

Since the previously mentioned graphs to be extracted are both labeled and weighted topological multi-graphs, then the proposed data structures in this case are as follows:

**a) Vertex list:** This list holds the data of every vertex in the graph; they are the vertex's label (numerical value) and degree.

**b) Non-looped branches list:** This list holds the data of all the non-looped branches in the graph.

**c) Looped branches list:** Since the multi-graphs may contain looped branches, thus it will be much more appropriate to put them in their own list.

For both non-looped and looped branches lists, the vertex's labels in which these branches are connected to shall be included in both of them.

The branch's orientation defines whether the branch is directed (true) or non-directed (false). It is to note that for the directed branches, the order of the branch's vertices always indicates the positive direction.

The name of the branch is composed from the nature of the component in which it is represented by this branch, plus the index of the component which differs between the components with the same nature. For example: if the branch represents a resistor with the index 1, then its name will be: R1.

According to rule n°1, the RLC components have an affinity to be impedances or admittances, so this affinity will be indicated as the category of the branch. Hence, every branch shall have a category name that is obtained from the association of the component's affinity, and an index to differ between components with the same category. For example: if G have the following components: R1, L1, C1 and C2, their category names are (respectively): Z1, Z2, Y1 and Y2.

The weight of the branch is the category of the component. It can be defined by both of the branch's category name and type name.

The previous mentioned data structures for the branches are made to store the data of each branch without to consider its connection with the rest of the branches, this connection gives a look at the topology of the graph. So to illustrate it, the adjacency-incidence list is defined.

**d) Adjacency-incidence list:** In this representation, the graph's data are represented by using linked lists to store the entire incident branches (looped branches are excluded) to each vertex in the graph. In the same process, the neighbored vertices to each actual vertex which is connected to those branches are stored as well. Typically to construct this list, all the branches of the graph -through the non-looped branches list- are swept (in coordination with the vertex list), and the adjacency-incidence list is updated. As a result, one can identify the adjacent branches to a specific branch, and in the same time one can identify all the incident branches to a specific node.

**Example**: Let us make the adjacency-incidence list of the following circuit (Fig.1).

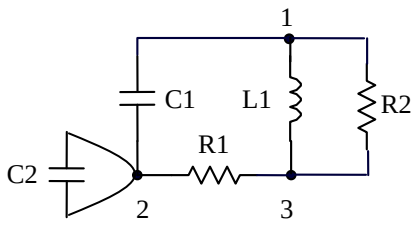



Fig. 1 Example of a circuit.

The graph associated to this circuit is called G.

This graph has one looped branch, so the looped branches list is as presented in Table 1.

Its Non-looped branches list as presented in Table 2.

Table 1: Looped branches list of graph G

| *Vertex Label* | *Component's Category Name* | *Component's Nature Name* |
|---|---|---|
| 2 | Y2 | C2 |

Table 2: Non-looped branches list of graph G

| *Branch's 1st Node* | *Branch's 2nd Node* | *Branch's Orientation* | *Branch's Nature Name* | *Branch's Category Name* |
|---|---|---|---|---|
| 1 | 2 | false | C1 | Y1 |
| 1 | 3 | false | L1 | Z3 |
| 2 | 3 | false | R1 | Z1 |
| 1 | 3 | false | R2 | Z2 |

This graph has three nodes, so the vertex list is as illustrated in Table 3.

Table 3: Vertex list of graph G

| *Vertex Label* | *Vertex Degree* |
|---|---|
| 1 | 3 |
| 2 | 2 |
| 3 | 3 |

Then, the Adjacency-incidence list is as presented in Table 4.

Table 4: Adjacency-incidence list of graph G

| *Vertex data* | | *Adjacency-incidence list data* |
|---|---|---|
| *Label* | *Degree* | *Adjacent Vertex → Incident Branch's Nature Name / Category Name* |
| 1 | 3 | 2 → C1/Y1 ; 3 → L1/Z3 ; 3 → R2/Z2 |
| 2 | 2 | 1 → C1/Y1 ; 3 → R1/Z1 |
| 3 | 3 | 1 → L1/Z3 ; 1 → R2/Z2 ; 2 → R1/Z1 |

Now, let us discuss the path's algorithm.

## 6. The weighted multi-graph's path algorithm WMGPA

The weighted multi-graph path algorithm WMGPA needs as a principal element for its development the adjacency-incidence list, because this list is simply a list of very elementary paths which contain only the beginning and ending nodes. From this remark, the WMGPA is as follows:

```
WMGPA ( BgnNd , EndNd, GRAPH )
{
    if ( BgnNd == EndNd ) {
        Printf ( "Error ! Beginning node shall be different than the ending node.") ;}
    else {
    TempPathCounter = Degree of BgnNd ;

    Creating path copies containg BgnNd in the temporary path array TempPath with a number equal to TempPathCounter ;

    for ( i = 0 ; i < TempPathCounter ; i++ )
    {
        Search in the Adjacency-incidence List of GRAPH for a match between data of the last node in TempPath[i];

        if ( last node in TempPath[i] == EndNd) {
            Save TempPath[i] in the PathList array ;
            Delete TempPath[i] ;
            Decrement TempPathCounter by 1 ;   }
        else if ( last node degree in TempPath[i] == 1 ) {
            Delete TempPath[i] ;
            Decrement TempPathCounter by 1 ;   }
    }

    for ( i = 0 ; i < TempPathCounter ; i++ )
    {
        Counter = Degree of the last node in TempPath [i];

        Create path copies from TempPath[i] , with a number equal to Counter and store them in TempPathAnalyzer ;

        Search in the Adjacency-incidence List for a match between data of the last node in TempPathAnalyzer ;

        Save the data of each neighboured element to the last node in the last position of each path stored in TempPathAnalyzer ;

        for ( j = 0 ; j < Counter ; j++ )
        {
            if ( last node in TempPathAnalyzer[j] is EndNd )
            {
                Save TempPathAnalyzer[j] in the PathList;
                Mark TempPathAnalyzer[j] as unwanted ;   }
            else if ( last node in TempPathAnalyzer[j] is repeated in it ) {
                Mark TempPathAnalyzer[j] as unwanted ;   }
            else if ( Degree of the last node in TempPathAnalyzer[j] == 1 ) {
                Mark TempPathAnalyzer[j] as unwanted ;   }
            else {
                Save TempPathAnalyzer [j] in TempPath ;
                Increment TempPathCounter by 1 ;
                Mark TempPathAnalyzer [j] as unwanted ;  }
        }
    }
  }
}
```

### 6.1 Description of the algorithm

This algorithm is intended to extract all the possible paths which start with the beginning vertex BgnNd and the ending vertex EndNd from a graph GRAPH.

As a first step, the algorithm processes the before mentioned vertices to make sure that they are not the same. Afterwards, it builds an array that stores the temporary paths called TempPath, and an array that stores the positively verified paths PathList.

There are two stages in this process:

As a first stage, TempPath stores BgnNd, then it starts to process -through the "adjacency-incidence" list- the neighboured nodes to this latter (without to forget their connecting branch). If the node to be processed is effectively EndNd, then this temporary path is verified as a complete path and it is stored in PathList (and eliminated from the temporary path array). If this is not the case, then if this node has a degree of one, then this temporary path is rejected due to the fact that this node can not pass to any other node anymore, then it can not pass to the ending node. If this condition is not correct, then this path is verified as temporary, and it may lead in the next "hop" to the ending node.

In the second stage, the last vertex is selected in order to get its neighboured vertices (and the branch connecting these two) by the mean of the adjacency-incidence list, then, another array arises which is TempPathAnalyzer, which presents a table that stores the established temporary "temporary paths" from the TempPath array, and the same procedure as in the first stage is set with an extra condition, which is the test of the pre-existence of this additional node in the temporary path, which means in other words that this node is repeated, and so it results in a rejection of this possibility. This process keeps looping until the temporary path array becomes void, at this state, all the possibilities have been considered and the algorithm finishes.

## 7. Application of the weighted multi-graph path algorithm in the DTM

The direct topological method may use WMGPA in many of its stages. One of its applications which is explicit and presents one of its most useful benefits is its ability to study the connectivity of the graph. As already mentioned in part III.B.4, if there is no way to find any path joining two nodes in a graph, then this graph is degenerated and as rule n°3 states, the determinant of a degenerated graph is null.
One of its implicit uses is when it is molded to become the cycle's algorithm, this algorithm can be used –in one of its applications– as a transfer cycle to calculate the topological transfer numerator according to the third theorem of this method (part IV.B).

## 8. Conclusion

In this paper, we proposed a new path algorithm that finds all the possible paths that start from a specific node to another in a graph. This algorithm is suited to the weighted multi-graphs; it can be considered as an extended DFS algorithm. It can be used to develop algorithm for DTM in order to analyze electrical circuits.

## References

[1] D. Narsingh, Graph theory with applications to engineering and computer science, Prentice-hall inc, 1974.
[2] M. C. Golumbic, Algorithmic graph theory and perfect graphs, Academic Press, 1980.
[3] R. Sedgwick, Algorithms in Java, Addison Wesley, third edition, 2003, part 5 : graph algorithms.
[4] D. Jungnickel, Graphs, networks and algorithms, Springer, third edition, 2007.
[5] H. Nagamochi and T. Ibaraki, Algorithmic Aspects of Graph Connectivity-Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2008.
[6] S. Hoang, Direct topological method, Course document, Telecommunications institute of Oran; Chapter4 : Theory of topological graph , Chapter 6 : The direct topological method for circuits analysis, page 177, 1984.
[7] S. S. Skeina and M. A. Revilla, Programming challenges, the programming contest training manual, springer, 2003.
[8] B. Courcelle, Introduction à la théorie des graphes : Définitions, applications et techniques de preuves, Université Bordeaux 1, LaBRI (CNRS UMR 5800), April 20th 2004.
[9] S. Hoang, "Direct topological method for analysis of networks without magnetic coupling", Arch Electrotech (Poland), 1974, 22(2) pp. 387-405.

**Abderrahmane Euldji** is a state engineer from the National Institute of Telecommunications and ICT (INT-TIC) since 2008. He is preparing for the magister degree in telecommunications, option: network systems and ICT. His current interest is computer science and algorithmic.

**Abderrahim Tienti** had his state engineer degree in telecommunications in 1983 from the National Institute of Telecommunications and ICT, a DEA in automatics from ENSIEG of INPG (Gronoble, France) in 1990 and his magister degree in robotics from University of Sciences and Technology of Oran (Algeria) in 1998. He is a lecturer at INT-TIC. His current interest is the analysis and synthesis of electrical circuits (especially in DTM), electronic switching systems, signaling system SS7, xDSL, ISDN. He has contributed in many papers in many national and international conferences.

**Amine Boudghene Stambouli** is a graduate of the University of Sciences and Technology of Oran (Algeria) in 1983. He received his master's degree in modern electronics (1985) and his PhD in optoelectronics (1989) at the University of Nottingham in England. He joined the University of Sciences and Technology of Oran in 1989. His studies started in the field of High Field Electroluminescence and optoelectronics and lately changed to environmental friendly production of energy. His research interests include at present: Photovoltaics, Fuel cells, hybrid systems, and

environment impacts. He is a full Professor of optoelectronics and material science for environment and energy applications at the Department of Electronics. Prof. Amine Boudghene Stambouli is United Nations consultant (Index 382958), member of many scientific and industrial organizations and director of several doctoral courses. He served as the head of the electronics department, vice rector of the university for almost two years and later the president of the scientific council of the electrical and electronics engineering faculty. Prof. Amine Boudghene Stambouli has been chairing five international conferences in the field of electrical engineering and was chairman at numerous sessions of international conferences. His studies are documented by 1 book, 3 polycopies and several papers mainly published on International Journals and on Proceedings of International and National Conferences. He is a reviewer and an Editorial Board Membership of several International Journals. He is actively collaborating with research group world wide (Algeria, Italy, Japan, France, USA, Turkey, England, Saudi Arabia, Jordan and Syria). He was Co-responsible, with Pr. Enrico travera, of the research team « Photovoltaics and fuel cells » between the university of Roma Tor Vergata and the university of Sciences and Technology of Oran. He was awarded the prise of the best publication of the year 2009, delivered by CDER and the ministry of higher education and scientific research of Algeria. He is Co-responsible (Algerian side) of the Sahara Solar Breeder (SSB) project along with Pr. Koinuma (Japan side). Founder of the Sahara Solar Breeder Foundation.

# Arabic Interface Analysis Based on Cultural Markers

Mohammadi Akheela Khanum[1], Shameem Fatima[2], Mousmi A.Chaurasia[3]
Information Technology Department, King Saud University
Riyadh, Kingdom of Saudi Arabia

**Abstract**

This study examines the Arabic interface design elements that are largely influenced by the cultural values. Cultural markers are examined in websites from educational, business, and media. Cultural values analysis is based on Geert Hofstede's cultural dimensions. The findings show that there are cultural markers which are largely influenced by the culture and that the Hofstede's score for Arab countries is partially supported by the website design components examined in this study. Moderate support was also found for the long term orientation, for which Hoftsede has no score.

***Keywords:*** *Cultural markers, Arabic website design, Hofstede's culture dimensions, Education websites, Media websites, Business websites***.**

## 1. Introduction

Human-Computer Interaction (HCI) is both an **art and a science**. The interdependence of a software system's functionality and its interface means that software designers cannot afford to favor one over the other. If the interface is well designed, it will allow the system's functionality to support the user's task. However, if the interface is inadequate, the functionality is obscured and users will have trouble accomplishing their task [1].With the growth and expansion of the Internet and World Wide Web, the number of non-English speaking users of Internet also is growing rapidly. For instance, the number of Arabic and Chinese speaking users has grown 1907.9% and 1087.7% respectively [2]. These figures are a clear indication that web connects various regions and various people across the world. The "one size fits all" formula no more holds for the web interface design. Therefore, the cultural context of user interface design has become an important issue to be considered while developing interfaces for different cultures. In this paper we discuss issues regarding the influence of culture on Arabic websites design. Arabic websites from three Arab countries (Saudi Arabia, UAE, and Kuwait) are taken for our analysis. We consider different educational websites, media websites and business websites, and analyze them according to Geert Hofstede's theory of culture dimensions [3]. The results of this study can be used by the web designers to design more localized interfaces for the Arab users.

## 2. Cultural Issues in Interface Design

### 2.1 Hofstede's cultural Values

There are many models of culture that are used by the researchers and practitioners which can help in studying and designing websites across cultures. Geert Hofstede's model is one of the most widely used models for studying the cross cultural challenges in the design of interfaces. Hofstede cultural dimensions are based on a large sample of employees from 40 countries from the large multinational IBM, whom he studied from 1960's, 70's and 80's. Hofstede's model comprises of five cultural dimensions. These dimensions comprise: *Power Distance*-the extent to which the less powerful members of organizations expect and accept that power is distributed unequally; *Individualism*-the extent to which individuals are integrated into groups; *Masculinity*-assertiveness and competitiveness versus modesty and caring; and *Uncertainty Avoidance*-intolerance for uncertainty and ambiguity. A fifth dimension, *Long-Term Orientation*-the degree of future orientation was added later on in 1982 when Hofstede expanded his model to include 10 more individual countries and three regions. As a group, Arab countries scored high on Power Distance (80), Uncertainty Avoidance (68), and Masculinity (52) dimensions, while scoring low on the Individualism (38) dimension. The only dimension that does not have any scores for these countries is the Long-/Short-Term Orientation dimension [3].

### 2.2 Related studies in non-Arab cultures

Researches in the past have focused on the effect of culture on the interface usability and influence of culture on the interface design. Some of the literature reviewed by us is as follows:

Radmila et al [4] in their study on developing UK and Korean cultural markers pointed to the general issues of the cross-cultural web design. To do this, they defined a checklist of relevant design elements which are supposed to be culturally specific design elements, called as the cultural markers [5]. The cultural markers they used included verbal attributes such as language and format (time, date, addresses, currency, printing format and size,

units of measurements), visual attributes which comprised of images, color, text, layout, and audiovisual attributes which includes sound, animation and 3D. The results indicate that some of the cultural markers were having same values across the UK and Korean websites and hence the authors concluded that such elements are not candidates for cultural markers and they may not influence the webpage design and usability.

Authors in [6] conducted a study on comparison of Malaysia and Britain local cultural values through the website analysis. They applied Hofstede's individualism/collectivism and power distance, and Hall's high/low-context cultural dimensions, to the various websites which included the university websites, tourism websites and bank websites. The result indicates that the cultural values presented in the local websites of the two varying cultures match the research of both Hofstede and Hall regarding the cultural differences between the countries.

Tong & Robertson[7] in their research on the political and cultural representation in Malaysian websites, adopted power distance from Hofstede's model of cultural analysis and Aaron Marcus's approach to multi-dimensional web-interface analysis to identify current representation of multicultural Malaysia. They used cultural marker's model (CMM) to investigate cultural inclusion. The results suggest that, it is not easy for designers to develop a sophisticated understanding of culturally sensitive visual interface design. Although there are many existing frameworks and theories, it is difficult for designers to identify the appropriate model for a particular multicultural society.

Kim & Kuljis [8] conducted a study on identification of elements that can be attributed to culture in the website design. They compared the South Korean and UK's charity websites based on Greet's Hofstede's theory of culture dimensions. The results show that there are some differences and preferences in the websites design that are mostly related to whether the websites employ multimedia and provide facilities for user input.

### 2.3 What do we know about Arab cultural markers?

Very fewer studies in the past have focused on the Arab interface design issues. Some of them are as follows:

Khusman et al. [9] proposed a model that includes cultural variables, which largely influence the user's acceptance behavior for the e-business websites in Arab countries. A field study was conducted in three main tourist sites of Jordan. Tourists from UK and Arab countries were used as the target group for this study. The questionnaire consists of questions related to the respondent's background and possible factors that may influence their acceptance and usage of e-business websites. The results suggests that e-business websites developed for low power distance, low uncertainty avoidance, high individualism and high masculinity cultures (like the Western cultures) are not optimally suited for Arab cultures which involves high power distance, high collectivism, low masculinity and high uncertainty avoidance. The analysis contradicts the results of Hofstede, which suggest that Arab cultures display a lower masculinity than western cultures.

Research by Aaron Marcus & Associates [10] in the year 2009, discusses issues regarding the influence of culture on Arabic websites. They analyzed three Arab countries, the Jordan, Egypt and the United Arab Emirates. The analysis was on the educational websites based on Geert Hofstede's theory of culture dimensions and Marcus (author) theory of user interface components. Marcus components include the metaphors, the mental model, navigation, interaction, and presentation styles of interface design. The results points out that Arabic websites need to consider some changes, such as, to add more representative pictures, more multimedia components, more links to the external websites, and more multilingual contents.

Study conducted by Khashman & Large [3] examine the design characteristics of government web interfaces from three Arab countries using Hofstede's cultural dimensions. Organizational and graphical elements from 30 ministry websites from Egypt, Lebanon and Saudi Arabia were examined using content analysis. Element frequency scores were correlated with Hofstede's dimensions and interpreted based mainly on the model developed by Marcus and Gould. The results suggest that Hofstede's model of culture does not fully reflect the design characteristics of Arabic interfaces.

## 3. Method

In order to understand the user interface requirements for any population, it is important to analyze the cultural influence on the acceptance of technologies. The primary aim of this research is therefore is to explore the cultural values of the Arab countries through analyzing their website designs. We used systematic analysis of cultural markers to find out how much the websites design conforms to the Hofstede's cultural dimensions. Also, we analyzed the cultural markers that are most prominent in Arab culture. We applied the list on a selection of 27 web pages from three countries in the Arabian Gulf region. These include Saudi Arabia, United Arab Emirates and Kuwait which are believed to have similar cultures. Most of the previous studies have focused on only one genre, either educational, Government or the tourist websites. Our perception is that the more the number of genres used, clearer would be the insight. Therefore, we chose Web pages from the following three genres (9 Web pages per

genre)

- Education
- News and Media
- Business

These genres are chosen on the basis that the websites for these genres are created by and for the people belonging to the local culture and consequently reflecting the socio-cultural and technological characteristics of their culture so as to be successful in providing the services to the target population. We selected the home pages for analysis, as they may contain many central elements of web design.

Table 1: List of Cultural Markers Used

| | | |
|---|---|---|
| ***COLOR*** | Background | |
| | Text | |
| | Link | Visited Link |
| | | Unvisited ink |
| | Menu Background | |
| | Menu Font | |
| | Picture | |
| | Logo | |
| ***LAYOUT*** | Menu Click | |
| | Menu Place | |
| | Scroll Bar | |
| | Page Orientation | |
| | Graphic / Image | |
| | Logo | |
| ***TEXT*** | Typeface | Title |
| | | Body |
| | | Menu |
| | Text Size | Title |
| | | Body |
| | | Menu |
| ***LANGUAGE*** | Native*(web site in native Arabic language only)* | |
| | Foreign (*non-Arabic interface only*) | |
| | Multiple (*bilingual Arabic/English or additional languages*) | |
| ***# OF LINKS*** | Internal | |
| | External | |
| | Total | |

## 3.1 Cultural Markers

The list of cultural markers that we used is listed in Table 1. Cultural markers were based on the components proposed by Barber and Badre [5]. Studies by Galdo [11], Fernandes [12] and Russo and Boor [13], shows that cultural factors such as the icons, color, symbols and language are essential elements to be considered while designing website. The web pages were analyzed based on five characteristic elements of web design which include Layout, Color, Text, Language, and Number of Links. These cultural markers have been used in several HCI studies which examine cultural usability [4, 5].

## 3.2 Stimuli

Variability in the sample being examined was considered to ensure that web pages were representative of the culture in the Arabic Gulf Region.

### 3.2.1 Education

Web pages for Education websites included nine organizations of higher education. These nine university web pages are categorized by Country and described in the figure below. For Saudi Arabia, web sites for three universities were selected. The first is the King Saud University, the first and the largest university in the Kingdom of Saudi Arabia. The second is King Fahad University of Petroleum and Mineral Sciences. Third is the King Abdulaziz University in the city of Jeddah in the Western province of the Kingdom. Fig. 1 shows the screenshots of the three education websites from Saudi Arabia

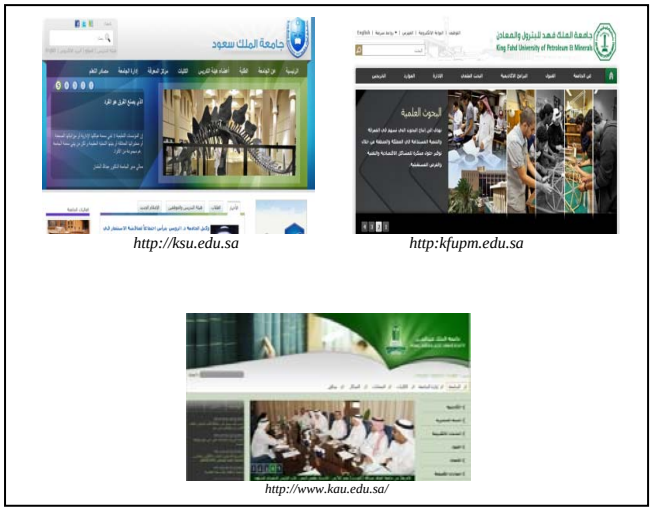
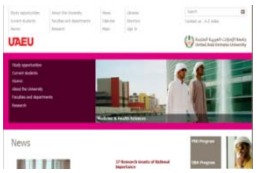



Fig. 1 Education Websites from Saudi Arabia

For the UAE websites, three universities were selected covering both private and public higher educations. They include the UAE University, the leading and pioneering educational institution in the region, the Zayed University in Dubai and Abu Dhabi and the Sharjah University, a leading institution for higher learning located in Sharjah. Fig.2 contains the screenshots of the education websites from UAE.



http://www.uaeu.ac.ae

http://www.zu.ac.ae

http://www.sharjah.ac.ae

Fig. 2 Education websites from UAE

For Kuwait, the three university websites comprising of the Kuwait University, the Global University of Science and Technology, and Kuwait Institute for medical specialization, are analysed for the cultural markers. Fig.3 depicts the homepages of the three education websites from Kuwait.

http:kuniv.edu.kw

http://www.gust.edu.kw


http://kims.org.kw

Fig. 3 Education websites from Kuwait

### 3.2.2 News and Media

From Saudi Arabia, three prominent and widely read newspaper's websites were analyzed. The Aljazirah and the Okaz Arabic newspaper were established in 1960, whereas the Asharq Alawsat was launched in London in 1978. Aljazirah is the first publication in Saudi Arabia. Okaz is the Arabic sister newspaper of Saudi Gazette. Asharq Alawsat, the first Arabic daily newspaper to execute satellite transmission for simultaneous printing in on four continents in 12 cities. Fig.4 displays the screenshots of the three media websites from Saudi Arabia.

http://www.al-jazirah.com

http://www.aawsat.com/

http://www.okaz.com.sa/new/Issues/20101220/index.htm

Fig. 4 Media websites from Saudi Arabia

From UAE, we considered three mostly read newspapers. The Alittihad is the oldest newspaper in UAE. Alkhaleej is a daily Arabic language newspaper published in Sharjah and established in 1970.Albayan was established in 1980 by the government of Dubai. The three media websites of UAE are displayed in Fig.5.

http://www.alittihad.ae/

http://www.alkhaleej.ae/

http://www.albayan.ae/

Fig. 5 Media websites from UAE

Three mostly read newspapers from Kuwait are considered for the analysis. They include the Alanba, Alqabas, and the Alrai. Fig.6 is the screenshots of three media websites from Kuwait.

### 3.2.3 Business

Nine Business organizations' websites where chosen for analysis. From Saudi Arabia three websites which include Saudi Aramco one of the leading oil and gas company, Saudi Basic Industries Corporation, leading manufacturers of chemical, fertilizer, Plastics and metal and Arab Petroleum Investments Corporation. Websites screenshots of three business organizations is shown in Fig.7.


http://www.alanba.com.kw/


http://www.alqabas.com.kw/


http://www.alraimedia.com/Alrai/

Fig. 6 Media websites from Kuwait


http://www.saudiaramco.com/Ar_sa/home.html


http://www.sabic.com/corporate/ar/default.asp


http://www.apicorp-arabia.com/html/cms

Fig. 7 Business websites from Saudi Arabia

Three websites from UAE includes the Abu Dhabi National Oil Company, which is a petrochemical company, the Crescent Petroleum and the Abu Dhabi Company for Onshore Oil operations. Fig.8 shows the screenshots of three business websites from UAE.


http://www.adnoc.ae/arabic/default.aspx


http://www.crescent.ae/ar/index-ar.html


http://www.adco.ae/home/index.aspx

Fig. 8 Business websites from UAE

From Kuwait we have selected the Kuwait National Petroleum company, the Kuwait petroleum Corporation and the Kuwait oil company. Fig.9 is the display of the business websites from Kuwait.


http://www.kockw.com/kocarabic/Pages/Default.aspx


http://www.knpc.com.kw/ar/Pages/default.aspx


http://www.kpc.com.kw/ar/default.aspx

Fig. 9 Business websites from Kuwait

## 4. Result

The analysis was conducted on these websites and 3 raters were used to examine the cultural markers.

### 4.1 Markers for color

Color is one of the strongest cultural markers. Color in our analysis refers to the colors of webpage background, typography, links pictures and logos. Different colors represent different meanings in different cultures, for example, Red color in China means happiness but danger in the US [5]. Therefore, use of color in interface design may have a greater impact on the user's satisfaction and expectations [7]. The cultural markers analysis of the result shows that white color, which symbolizes purity and peace in the Arabian culture, has been used prominently as background color (85%) as well as the menu font color (48%). The color blue, which has been suggested to indicate protection in Arabian culture [15], has been used as the visited and unvisited links color (55%) and also as the menu background color (26%) in most of the web pages that we analyzed. Logo, which represents the historical background of the country, can be found on each web site. Multiple colors are usually used for the logo with green and blue dominating. The color markers results summary is depicted in Table 2.

Table 2: Result summary for Color Markers

<table>
<tr><th colspan="7">Color</th></tr>
<tr><th rowspan="2">Background</th><th rowspan="2">Text</th><th colspan="2">Link</th><th rowspan="2">Menu</th><th rowspan="2">Menu font</th><th rowspan="2">Image</th><th rowspan="2">Logo</th></tr>
<tr><th>Visited</th><th>Unvisited</th></tr>
<tr><td>White=23</td><td>Gray=14</td><td>Blue=14</td><td>Blue=15</td><td>Blue=7</td><td>White=13</td><td rowspan="10">Multiple=27</td><td>Multiple=13</td></tr>
<tr><td>Maroon=1</td><td>Black=6</td><td>Green=4</td><td>Green=4</td><td>White=5</td><td>Blue=6</td><td>Blue=5</td></tr>
<tr><td>Green=1</td><td>White=2</td><td>Gray=3</td><td>Gray=3</td><td>Gray=4</td><td>Gray=3</td><td>Green=5</td></tr>
<tr><td rowspan="6">Blue=2</td><td>Green=2</td><td>Purple=1</td><td>Red=2</td><td>Green=3</td><td>Green=2</td><td>Black=3</td></tr>
<tr><td>Multiple=2</td><td>Red=1</td><td>Maroon=1</td><td>Black=2</td><td>Black=2</td><td rowspan="5">White=1</td></tr>
<tr><td rowspan="4">Blue=1</td><td>Maroon=1</td><td>Orange=1</td><td>Orange=1</td><td rowspan="4">Red=1</td></tr>
<tr><td>Orange=1</td><td rowspan="3">Multiple=1</td><td>Maroon=1</td></tr>
<tr><td rowspan="2">Multiple=1</td><td>Red=1</td></tr>
<tr><td>Yellow=1</td></tr>
<tr><td>Multiple=1</td></tr>
</table>

## 4.2 Markers for layout

According to Yu and Roh [14],” appropriate design layout provides web visitors with a contextual and structural model for understanding and accessing information”. Layout of the web pages which includes the menu placement, the place of the scroll bar, the page orientation and the place of the logo, identifies the cultural preferences. All the surveyed websites are vertically oriented with a side scroll bar. Static menus prominently placed on the top can be found on most (52%) of the websites. Some websites also have menu placed both on the top as well as on the left and right side of the page. Images on the websites are placed prominently on the top; some images are also found on different parts of the webpage. Images show the official logos, official buildings, students, and official authorities (deans, chairmen, and the founder). The result of layout analysis is shown in Table 3.

Table 3: Result summary for Layout Markers

<table>
<tr><th colspan="6">Layout</th></tr>
<tr><th>Menu click</th><th>Menu place</th><th>Scroll bar</th><th>Page orientation</th><th>Graphic/Image</th><th>Logo</th></tr>
<tr><td>Static=14</td><td>Top horizontal=18</td><td rowspan="6">Vertical=27</td><td rowspan="6">Vertical=27</td><td>Top=10</td><td>Top Right=21</td></tr>
<tr><td>Drop Down when mouse hovers= 9</td><td>Right vertical=4</td><td>Everywhere=9</td><td>Top Left=5</td></tr>
<tr><td rowspan="4">Drop Down = 3</td><td>Center=2</td><td>Center=1</td><td rowspan="4">Center=1</td></tr>
<tr><td>Multiple=2</td><td>left side=5</td></tr>
<tr><td rowspan="2">Vertical Left=1</td><td>Right side=1</td></tr>
<tr><td>No graphic/image=1</td></tr>
</table>

## 4.3 Markers for text

Text is another important cultural marker which strongly represents the cultural preferences. The type of fonts used in the web pages has an impact on its usability. In our survey of the websites, we found the font types Tahoma and Arial are preferred fonts for the title, body and menu. The text size for the body, menu and title ranges from 12px to 15px.The text analysis results are depicted in Table 4.

Table 4: Result summary for Text Markers

| ***Text*** | | | | | |
|---|---|---|---|---|---|
| ***Type face*** | | | ***Text size*** | | |
| ***Text*** | ***Body*** | ***Menu*** | ***Text*** | ***Body*** | ***Menu*** |
| Image=19 | Tahoma=7 | Image=7 | Image=19 | 11px=3 | Image=7 |
| Normal=5 | Arial=5 | Arial=4 | Medium=4 | 12px=7 | 10px=2 |
| Tahoma Bold=1 | Multiple=4 | Multiple=4 | 32px=1 | 13px=3 | 11px=1 |
| Italic Bold=1 | Arabic Transparent=4 | Tahoma=3 | 15px=2 | 14px=2 | 12px=2 |
| Arial=1 | Simplified Arabic=2 | Normal=3 | 12px=1 | 15px=4 | 13px=4 |
| | Normal=2 | Transparent Arabic=2 | | 16px=3 | 14px=3 |
| | Traditional Arabic=1 | Bold=1 | | 18px=1 | 15px=3 |
| | Image=1 | Myriad Pro=1 | | 19px=1 | 16px=2 |
| | Verdana=1 | Times New Roman=1 | | 21px=1 | 20px=1 |
| | | Simplified Arabic=1 | | Image=1 | 26px=1 |
| | | | | | 29px=1 |

## 4.4 Markers for languages and links

The most distinctive cultural symbol is language. In everyday usage, Arabic is most commonly shared language for Arab society although people have different local dialects and culture, the learning of English and its use is common in everyday life. The bilingual websites are developed probably due to the fact that there are minorities and many foreign workers in these countries who do not speak Arabic. The language preferences have been tested based on native (native Arabic language only), foreign (non-Arabic interface only) and multiple (bilingual Arabic/English or additional languages). The result analysis shows that most of the websites are bilingual (52%), 30% of the websites have only Arabic interfaces with a majority being the media websites. Very few (18%) have only English interface. Apart from this, a fewer websites also have interfaces in other languages such as Spanish and French. The websites having only English interface could pose a barrier for the Arabic speakers. Links are used everywhere, in navigation, in banners, and in graphics. We evaluated the total number of links (sum of internal and external links) for each of the website using the link counter available on [16]. We considered internal links as those links which will open in the same window and the external links will open in a new window. 93% of the total links open in the same browser window and only 7% of the links open in the new browser window. These figures are indicative of the sequential way of Arabic culture preference in solving the problems. The result summary for languages and links can be found in Table 5.

Table 5: Result summary for Languages and Links

| ***Language*** | | | ***# of links*** | | |
|---|---|---|---|---|---|
| ***Native*** | ***Foreign*** | ***Multiple*** | ***Internal*** | ***External*** | ***Total*** |
| 8 | 5 | 14 | Min=0 | Min=0 | Min=0 |
| | | | Max=303 | Max=27 | Max=307 |
| | | | Total=2369 | Total=178 | Total=2547 |

## 5. Discussions

In this study, we examined cultural markers in a selected sample of Arabic web sites. The analysis revealed patterns of usage of cultural markers and in this section we describe how these findings relate to Hofstede's dimensions. Hofstede dimension says Arab culture has high power distance (80). This claim is supported by our findings which include the images of leaders (63%), images of official buildings (33%), and official logos (100%).Arab countries scoreless (38) in individualism and collectivism according to Hofstede dimension. Our findings which include the group pictures, lesser authentication passwords, supports Hofstede's claim. We found most websites have the pictures of males. Few of them also have the pictures of both males and females together, but females are covered in Abaya (traditional veil).This partially favors the Hofsetde's score for Arab countries, which says Arab countries have a score of 52 in masculinity versus femininity. The presence of simple menus, and detailed information supports Hofstede's claim that Arab countries have high (68) uncertainty avoidance. Hofstede dimension has no score for the long term versus short term orientation for Arab countries. We found 96% of the total websites surveyed have a search engine and 48% have site maps. Alumni links where found in fifty percent of the university websites. Moderate support for long term orientation was found based on site maps and alumni links. Our findings suggest short term orientation, however further research examining this factor/dimension is needed.

## 6. Conclusion

This study examined cultural markers of Arabic websites. Findings indicate that there are cultural markers which are largely influenced by the culture and the Hofstede's score for Arab countries is partially supported by the website design components examined in this study. Hofstede has no score for long term versus short term orientation, for which our analysis resulted in moderate support for long term orientation.

**Acknowledgment**

This research project was supported by a grant from the Research Center of the Female colleges for Medical and Scientific studies in King Saud University. We are immensely thankful to Dr. Areej Al Wabil, without whose help this work would have been impossible.

## References

[1] J. F. Dix, G. D. Abowd & R. Beale, "Human-Computer Interaction," 3 edn, Addison-Wesley Pearson Education, London. 2004.

[2] Miniwatts Marketing Group. Internet growth and stats. 2000-2010. Miniwatts Marketing Group. [Online]. Available: http://www.internetworldstats.com

[3] N. Khashman & A. Large, Measuring Cultural Markers in Arabic Government Websites Using Hofstede's Cultural Dimensions. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Vol .6770, 2011.

[4] R. Juric, I. Kim & J. Kuljis, Cross cultural web design: an experience of developing UK and Korean cultural markers, *Proc. 25th International ConferenceInformation Technology Interfaces ITI* 2003, Cavtat, Croatia. pp 309-313.

[5] W. Barber & A. N. Badre, Culturability: The merging of culture and usability, *Proc. 4th Conference on Human Factors and the Web*, Baskin, Ridge, New Jersey, 1998.

[6] T. Ahmed, H. Mouratidis & D. Preston, "Website Design and Localisation: A Comparison of Malaysia and Britain," *International Journal of Cyber Society and Education*, Vol. 1, No. 1, March 2008, pp 3-16.

[7] M. C. Tong, & K. Robertson, "Political and Cultural Representation in Malaysian Websites," *International Journal of Design*, Vol.2 No.2, August 2008, pp, 67-79,.

[8] I.Kim, & J. Kuljis, "Manifestations of culture in website design," *Journal of Computing and Information Technology*, Vol.18, No.2, 2010, pp125-132.

[9] S. Khushman, A. Todman, & S. Amin, "The Relationship between Culture and E-business Acceptance in Arab Countries," *2009 Second International Conference on Developments in eSystems Engineering (DESE*), 14-16 Dec. 2009, pp.454-459.

[10] A.Marcus , "The Impact of Culture on the Design of Arabic Websites," *Proc. HCI International Conference*, San Diego, CA, 19-24 July 2009.

[11] D. Galdo, & J. Nielson, International user interfaces, New York: John Willey and Sons, 1996.

[12] T. Fernandes, Global interface design. London: Academic Press.1994.

[13] P.Russo, & S.Boor, "How fluent is your interface?Designing for international users," *In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factorsin Computing Systems*, New York: ACM.1993, pp. 342-347.

[14] B .Yu, & S. Roh, "The effects of menu design on information-seeking performance and user's attitude on the web," *Journal of the American Society for InformationScience and Technology*, Vol. 53 No.11, 2002, pp 923-933.

[15] [Online]Available: http://www.colourlovers.com/blog/2007/09/08/colors-of-religion-islam/

[16] Link counters [Online]. Available: http://linkcounter.submitexpress.com

# Hybrid framework for mitigating illegitimate Peer Nodes in Multimedia file sharing in P2P

Mr. Ramesh Shahabadkar[a], Dr. Ramachandra V. Pujeri[b]
[a]KResearch Scholar, Anna University, India,
[b]KGiSL Institute of Technology, Coimbatore, India

**Abstract: Peer to Peer network is one of the frequently used application in terms of file sharing over a global large network. In such types of network, there can exists an illegitimate peer node who will attempt to have an unauthorized access to premium digital content. As it is very difficult to catch hold of the intruder or the illegal client inside the network, the proposed system will at least attempt to prevent access to the digital content available. The proposed system is focused on multimedia file sharing where the protocol will be designed in such a way that whenever any unauthorized peer node will attempt to download any premium digital content, the proposed model will assign a poison chunk of data to be forwarded to the illegitimate client. This phenomenon will result in exponential increase of download time rending discouragement to download the same file by the illegal downloader. While the protocol assures the secure delivery of the cleaner chunk of data to the legitimate client.**

**Keywords**: Peer to peer, security,. content poisoning.

## I. INTRODUCTION

The peer-to-peer multimedia-sharing network is recent trend in delivering large amount of file to enormous number of participants [1]. The unauthenticated users are punished by means of poisoning the network, which raises immense legal issues in the field of P2P network. The prominent reason of the illegitimate multimedia file sharing are found as peer nodes which overlook the various rights as well as attempt to mitigate the various illegitimate peer nodes. Conventional content delivery networks [2] will use a huge quantity of the proxy digital content servers spread over globally distributed WANs. The content distributors need to replicate or cache contents on many servers. The bandwidth demand and resources needed to maintain these CDNs are very expensive. A P2P content network significantly reduces the distribution cost [3], since many content servers are eliminated and open networks are used. P2P networks improve the content availability, as any peer can serve as a content provider. P2P networks are desired to be scalable, because more peers or providers lead to faster content delivery.

The legitimate clients are considered as the entity in the network which obeys the digital rights management and other restrictions. Pirates are peers attempting to download some content file without paying or authorization. The colluders are those paid clients who share the contents with pirates. We use identity-based signatures (IBS) [4] to secure file indices. IBS offers the same level of security as PKI-based signatures with much less overhead. We apply discriminatory content poisoning against pirates. We focus on protection of decentralized P2P content networks. Protecting centralized P2P networks like Napster or mp3.com is much simpler than the scheme we proposed.

Our goal is to stop collusive piracy within the boundary of a P2P content delivery network. Our protection scheme works nicely in a P2P network environment. The scheme cannot stop randomized piracy in open Internet using Email attachment or any other means to spread copyrighted contents, illegally. Randomized piracy is beyond the scope of this study.

The remainder of this paper is organized as follows. Section II presents P2P content poisoning followed by related works in Section III. The proposed approach is discussed in Section IV. Section V will highlight the system model deployed in this research work followed by Implementation in Section VI and Simulation result in Section VII and finally Section VIII will conclude the research proposal.

## II. P2P CONTENT POISONONG

Currently, content owners only distribute free-of-charge files such as product demo, freeware and open-source software via P2P file-sharing networks. They do not distribute content to paid customers via P2P networks due to the high risk of copyright infringement. If any of the paid content is found in P2P networks, they are pirated. In that case, content owners will use content poisoning to disrupt the illegal file distribution. The disruptions appear as failed distribution, corrupted contents, or repeated frustrations on the part of unauthorized downloading clients. The concept of content poisoning in such a P2P system is illustrated in Fig.1.

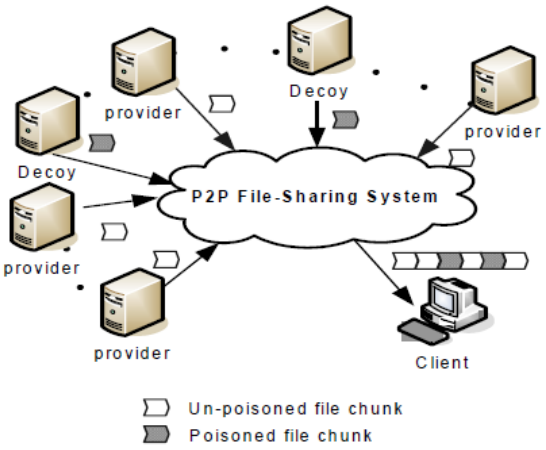



Figure 1: Content poisoning in a P2P file-sharing System, where providers generate clean file chunks and the decoys generate poisoned chunks

The illegal content providers on a P2P network are the copyright violators. They share clean file chunks to other peers. The content owners deploy a small number of decoys to poison the copyrighted files at subdivided chunk level. Unlike providers, decoys send out poisoned chunks in response to client's download requests. In the sequel, we use the term peer to refer to either a provider or a decoy, because both are not distinguishable by clients. The number of decoys divided by the total peer count is the decoy density.

For a real example in August 2006, we discovered a piracy case, by which a brand-name word-processing package was illegally distributed over a well-publicized P2P network. We hide the real names and fileID involved in order to protect all the parties involved. We refer the software file by a fake name: XYZ Pro. This file was measured to have 643.63 MB, subdivided into 3571 chunks. The file is identified by a fake fileID =XYZ1234567890Pro. When requesting the file, 56 peers on the P2P network responded, among them 8 peers are immediately available for file transfer. By deploying decoys into P2P file-sharing systems, owners want to discourage the download of copyrighted files by all clients, who did not subscribe the requested file. The major design issue of content poisoning lies in cost-effective deployment of decoys. Effective poisoning should use only a few decoys. Excessive decoying is unnecessary and cost-prohibitive. Content owner should also anticipate that those clients download pirated content off the P2P network will apply various techniques to resist poisoning.

***Definition of Content Poisoning Effect***

Many technical factors affect the download performance of a P2P file-sharing system, such as network topology, routing scheme, traffic congestions over the network, geographical locations of peers, etc. In order to single out and focus on the effect directly caused by content poisoning, we need to filter out all other effects. Let S be the actual file size (in Mega bytes) and D be the total number of bytes downloaded. We define the poisoning effect by:

$$\text{Poisoning Effect} = 1 - S / D$$

Poisoning Effect isolates the download effort wasted due the existence of decoys providing poisoned chunks. Its value represents the portion of downloaded bytes that are wasted due to the existence of decoys in a P2P file-sharing system. For example, in an ideal P2P file-sharing system where no decoy was present, we have S = D, meaning the client received exactly the same amount of bytes as the actual size of the file. Thus, the poisoning effect becomes zero. The poisoning effect approaches 100%, if D becomes extremely large, meaning most download requests failed. The content owner does not care of the copyright violator's interest. Their goal is to achieve higher poisoning effect. For example, 0.9 poisoning effect implies that the client must download 10 copies of the same file in order to retrieve an un-poisoned version. Since the owner's cost is directly related to the number of decoys deployed, a cost effective approach is to deploy just enough decoys.

## III. RELATED WORK

Nicolas e.t. al [5] conduct a measurement study of content availability in four of the most popular peer-to-peer file sharing networks, in the absence of poisoning, and then simulate different poisoning strategies on the measured data to evaluate their potential impact.

Ruichuan e.t. al [6] proposed poisoning-resistant security framework for P2P content sharing systems which is able to defend against the content poisoning attack effectively and efficiently.

Stephanos [7] study current peer-to-peer systems and infrastructure technologies in terms of their distributed object location and routing mechanisms, their approach to content replication, caching and migration, their support for encryption, access control, authentication and identity, anonymity, deniability, accountability and reputation, and their use of resource trading and management schemes.

Dan [8] has proposed a fully functional identity-based encryption scheme (IBE). The scheme has chosen ciphertext security in the random oracle model assuming a variant of the computational Diffie-Hellman problem.

Lei Guo [9] has presented a novel and efficient design of a scalable and reliable media proxy system supported by P2P networks. This system is called PROP abbreviated from our technical theme of "collaborating and coordinating PROxy and its P2P clients" with an objective is to address both scalability and reliability issues of streaming media delivery in a cost-effective way.

Ernesto Damiani [10] propose a self-regulating system where the P2P network is used to implement a robust reputation mechanism. Dumitriu [11] consider the file targeted attacks in current use in the Internet, and we introduce a new class of p2p-network-targeted attacks.

Tom [12] propose a P2P protocol that integrates the functions of identification, tracking and sharing of music with those of licensing, monitoring and payment.

Balachander [13] discuss how CDNs are commonly used on the Web and define a methodology to study how well they perform. A performance study was conducted over a period of months on a set of CDN companies employing the techniques of DNS redirection and URL rewriting to balance load among their servers.

Stefan [14] examines content delivery from the point of view of four content delivery systems: HTTP web traffic, the Akamai content delivery network, and Kazaa and Gnutella peer-to-peer file sharing traffic. Pablo Rodriguez [15] has discussed that although P2P technology has been widely associated with the distribution of pirated content and has been subject to a barrage of attacks (e.g. DoS, spoofing and content pollution), and there are ways to decrease the risks associated with distributing content using P2P technology.

Matthew [16] has analyzed the cost of the implementing the redundant task allocation in order to prevent illegal cases over internet. Kevin Walsh [AR] employed a novel voter correlation scheme to weigh the opinions of peers, which gives rise to favorable incentives and system dynamics by presenting simulation results indicating that system is scalable, efficient, and robust.

Runfang Zhou [17] proposes a gossip-based reputation system (GossipTrust) for fast aggregation of global reputation scores which leverages a Bloom filter based scheme for efficient score ranking. GossipTrust does not require any secure hashing or fast lookup mechanism, thus is applicable to both unstructured and structured P2P networks.

Currently, many reputation models have been proposed to address the problem of content poisoning in P2P content sharing systems. In general, these reputation models can be grouped into three categories: peer-based models, object based models and hybrid models. In peer-based reputation models, e.g., EigenTrust, PeerTrust and Scrubber, genuine users collectively identify content poisoners by computing a reputation score for each user, and then isolate these poisoners from the system. However, the studies in old research work implied that these peer-based models are insufficient to defend against the poisoning attack.

To address such problems we are going to propose proactive content poisoning system In contrast, our scheme detects unpaid pirates and use discriminatory content poisoning to deter on-line piracy. Legitimate clients can still enjoy the flexibility and convenience provided by open P2P networks. Our scheme stops pirates from illegal download of copyrighted files, even at the presence of many colluding peers. We developed a reputation-based method to detect peer collusion in piracy process.

## IV. PROPOSED SYSTEM

The aim of the research work is to design such a protocol that should consist of reduced delivery cost, maximized digital content availability and copyright compliance while exploring P2P network resources. The proposed scheme should also be capable enough to mitigate the collusive piracy within the peer to peer content delivery network. The main objectives of the proposed research work can be briefed as following:

- Error-free identification of colluders and pirates.
- Creation of secure protocol for forwarding poisoned chunk of data to illegitimate client in network.
- Creation of protocol for preventing colluders to come in contact with legitimate clients in the network.

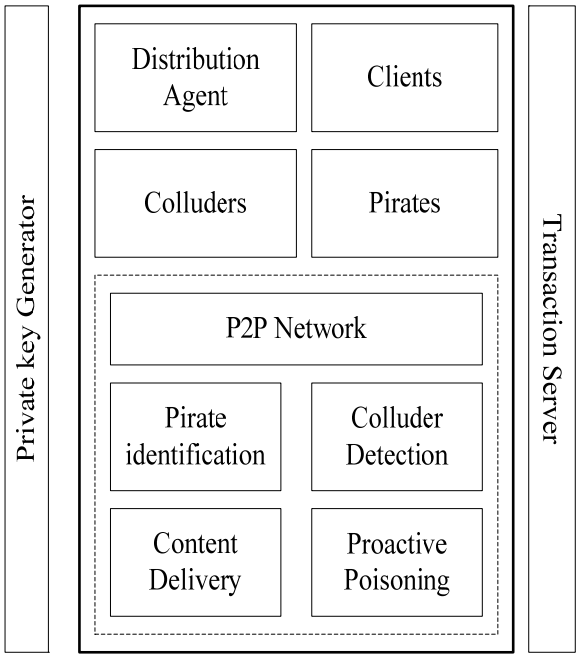



Fig.2 Overall Architecture of the proposed model

Fig 2 shows the overall architecture of the proposed "proactive content poisoning" system where the main components will be Distribution agent, Clients, Colluders, and Pirates. The next phase of architecture will consist of identifying the pirates and colluders, content delivery, and proactive poisoning. The technique used will be to initiate a design of methodology for authenticating peer with IP address and definitely port number. These entities will evaluate the component peers when the download or the upload of the digital content takes place where each of the peers will attempt to identify the illegitimate peers. The proposed approach will seek authentication for the peer looking to download or upload the file and in case the authentication fails for showing existence of illegitimate client in the network, our proposed system model will direct poisoned chunks to the illegitimate peer. The proposed technique can be suitably used in order to identify colluders. The flow of the proposed model is shown in Fig 3.

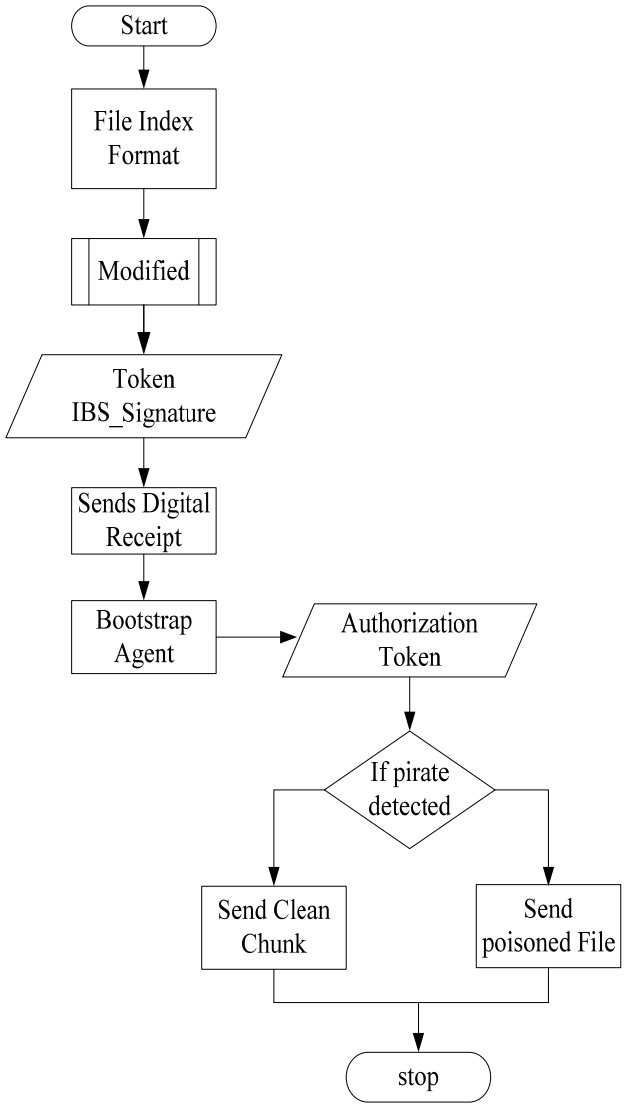



Fig.2 Proposed flow of system model

The above fig.3 shows the pro process flow diagram for our proposed system model of proactive content poisoning. The proposed technique uses the concept of forwarding poisoned chunks to discourage the illegitimate peer existing in the P2P network, where the restricted poison identification potential is exploited and enforce the pirate to reject the clean

chunks downloaded along with the poisoned chunks. Creation of such technique will induce the highly increased download time for the unauthorized client thereby discourage their interest and allow secure transmission to the service provider to legitimate client.

## V. SYSTEM MODEL

The proposed system is designed with 4 categories of peer as:

- Clients: This type of the peer are normally the authorized peer within the network
- Colluders: This type of the peers are paid peers who are sharing contents with other legitimate peers
- Distribution agents: These are the set of trusted peers which are controlled by the digital content owners for file distribution.
- Pirates: These are unpaid or unauthorized clients who are always at lookout of some unauthorized downloads of premium contents.

The proposed design also includes a transaction server (see fig 2.) which is assumed to be managing the buying and billing related issues. The transaction server will receive the request from the clients in the network. The communications among the peers are secured by configuring a private key generator in order to generate set of private keys with Identity-Based Signatures (IBS). The private key generator also acts as certificate authority in various secure communications. When the peers will attempt to get themselves connected in a network, the system will deploy the key generator and transaction server. Along with invocation of IBS, the peer communication will not be depended on any public key specially designed as their identity itself will serve as the key of public type.

The system model will initiate with colluders, paid clients, as well as pirate without indexing for their identification. The proposed prototype system model is designed to quantize them involuntarily. The model also consists of a bootstrap unit as an input point of entry which is chosen from one of the distribution agent. The existing network designs involuntarily declare its identification without any type of identification. In order to deal with this, the system model uses port number and IP address instead to any other types of authentication parameters that can be possible used in the same set of the network. A complete connection with the peer is only ensured when it is reachable via any listening port on its host.

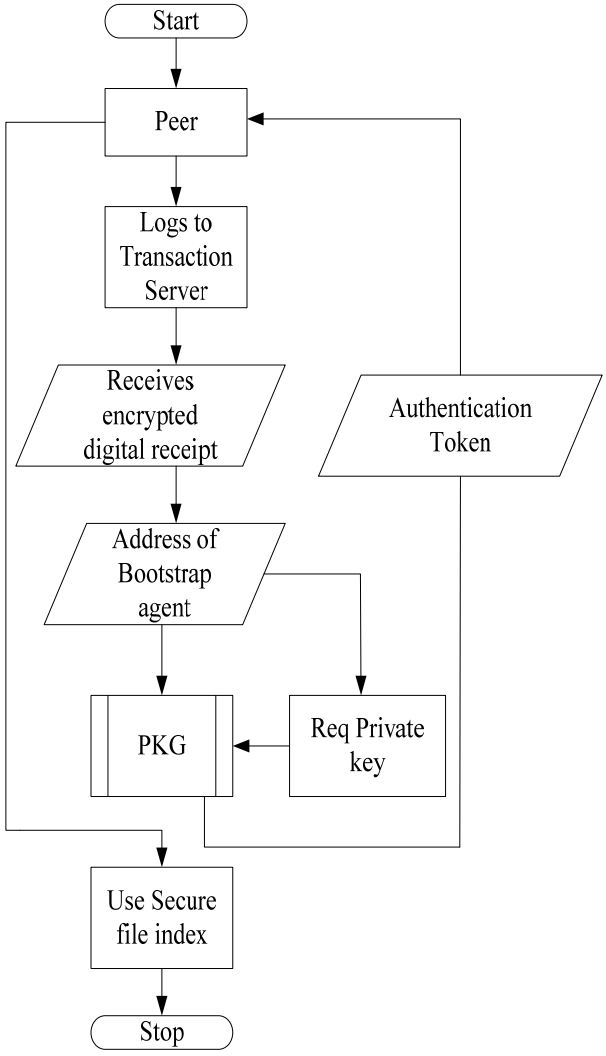



Fig 4. Peer Node Participation Process

The above Fig. 4 shows process flow diagram for peer participating process where it can be understood that each legitimate peer has a valid token (use indication). The token (indication) will be have possession of validity for a very minimized instant of time in order for the peer components to update in constant time duration. The file-index format is modified in order to include a token (indication) and signature, which will be used by the peer nodes for securing the download permission. The system model will use the identification of the listening port as valid identification of the peer. The model also assumes that every peer node posses a better configured listening ports. It was also found that the

majority of the peer to peer clients will linked themselves to the World Wide Web using their home network where the fundamental standard will be to deploy network address translation equipment for sending forward incoming ports. The issue surfaces when a large quantity of the peer is behind a single network address translation are used. The public key to be used in the peer's end will be considered as end point address for which there is absolutely no requirement of encoding the contents of the file thereby minimizing the feasibility of network overhead. The model deploys the bootstrap unit in order to forward the incoming request. The client module will be not disclosed about the identity of the all units apart from bootstrap unit, which prevents the malicious peer node to blacklist or initiate an attack on the distribution agent.

## VI. IMPLEMENTATION

The implementation of the proposed research work is carried out in 32 bit windows OS of 1.8 GHz with dual core processor. The programming is carried out in java platform. Although it is a very tough assignment to perform this experiment for real-time peer to peer network for evaluating the copyright infringement. The proposed simulation work is carried out in three phases:

- Estimation of the chunk poisoning rate.
- Estimation of the download time
- Comparative analysis of resistance and overhead in fortification information gather.

The simulation environment of the proposed system is as shown below in fig.5.

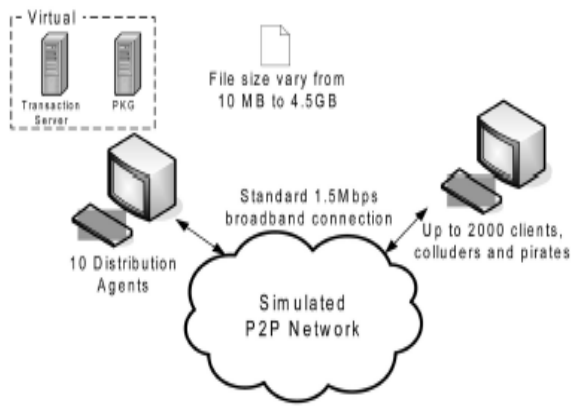



Fig.5 Simulation Scenario considered

The proposed framework will perform in three levels again:

- The top level will work towards emulating the peer to peer transmission.
- The intermediate level will be used for emulating the patterns of the considered peer types (distribution agents, legitimate clients, colluders, illegal pirates.)
- The bottom level will be allocated for data aggregation and information updating.

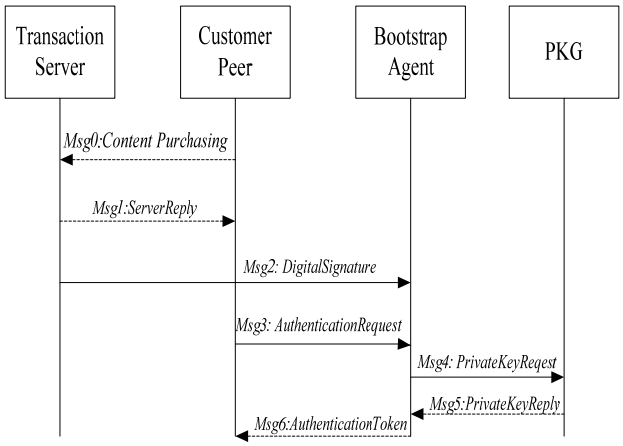



Fig.6. Sequence diagram of the simulation considered

As shown in Fig.6.the simulation is conducted in the same procedure. When a peer node will participate in the network, it has to initially access the transaction server in order to complete the payment process for possessing the digital content. Once the transaction is completed, a digital acceptance slip consisting of the basic information of the content and identification of the clients will be in possession of the client. This digital acceptance slip will be lock in such a procedure that only the legitimate owner of the digital content will be able to unlock it. The address of bootstrap unit is possessed by the authorized client. The newly created digital acceptance slip will be considered for authenticating the newly participating client with bootstrap unit. As the bootstrap unit is configured by the actual owner of the digital content, therefore it unlocks the acceptance slip and authenticate its verification parameters. The transaction server will assign a session key for securing the privacy in the communication channel. Design of a legitimate token takes place when the bootstrap unit request key from the key generator. The existence of the pirates are scrutinized by the peers by evaluating the legitimacy of the supplementary signatures in the index of the files. This scheme of security is deployed by the legitimate peer nodes to distribute clean digital

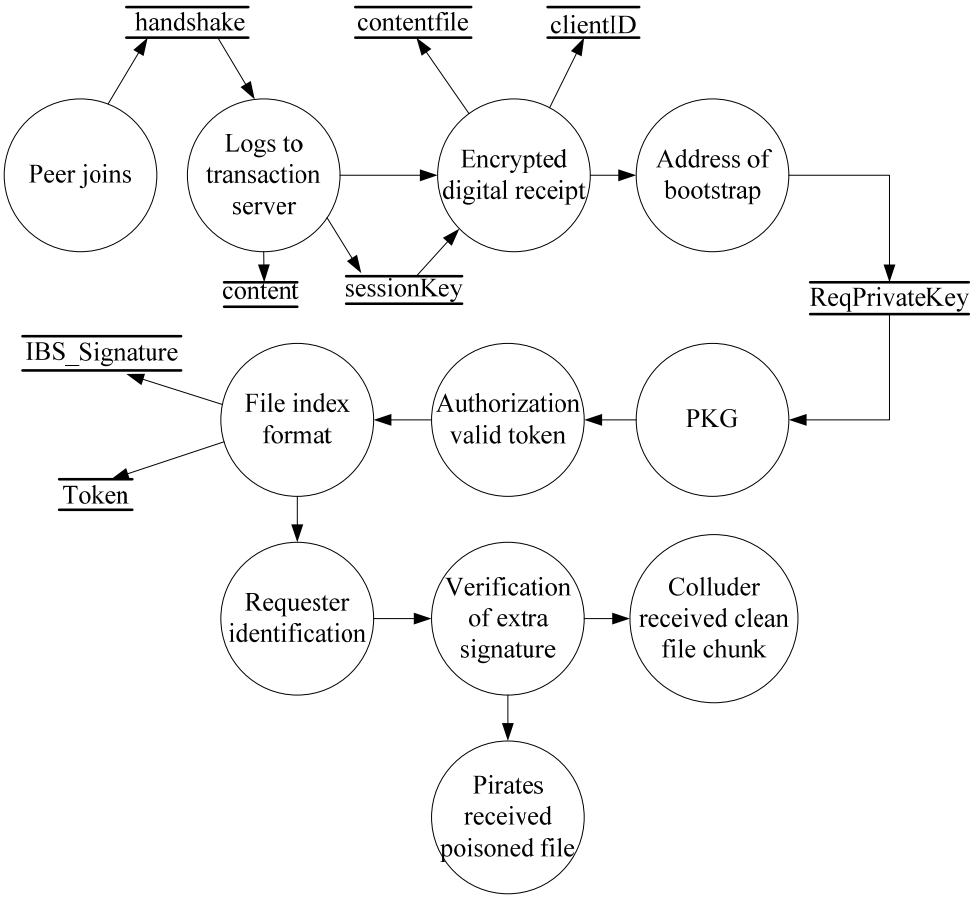



Fig.7.Data Flow Diagram of the proposed model

contents explicitly among the peer nodes and use the forwarding of the digital content poisoning procedure to the unauthorized clients in peer to peer network. The considered tokens which are time-stamped makes sure that identified colluders should not be able to posses newly generated token once the old token perishes.

## VII. SIMULATION RESULTS

The simulation is conducted in Java which estimates the shortest feasible download time by a pirate or by any normal clients. Practically, the download time should be highly increased for the case of unauthorized clients present in the network. The simulation is initiated by considering distributing agent to have a possession of clean chunks. The proposed simulation test bed has consideration of various challenging scenarios of piracy for understand the efficiency of the framework designed.

Fig 8 shows the simulation results considering average path length, network diameter, and algorithm for identity based signature and non-indexed system.

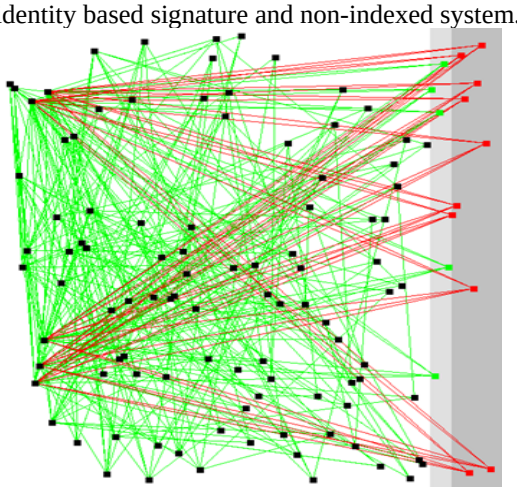


Fig 8. Simulation output showing peers

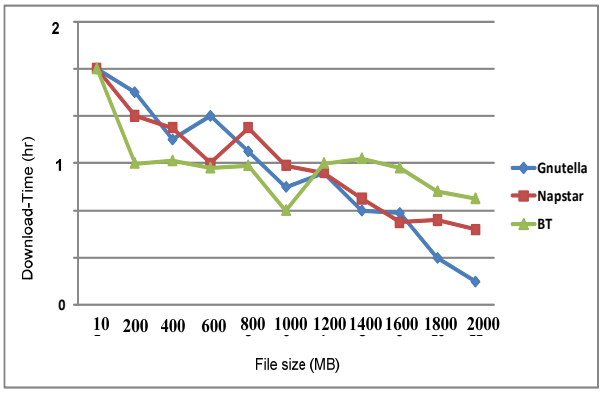



Fig. 9. Simulation result for authorized clients

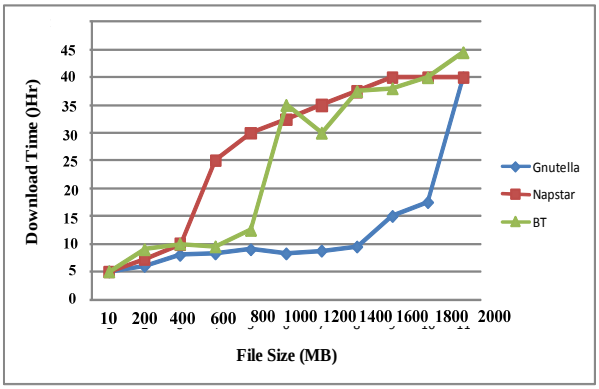



Fig 10. Simulation result for un-authorized clients

Any frequently used P2P protocol like BitTorrent or NapStar could be used for the purpose of simulation. The distribution is evaluated for 2 GB of a multimedia file downloaded from youtube.com. The proposed system is also evaluated with the other different size of the file which have different contents (other than multimedia). Understanding the security implementation of the proposed protocol, it has been seen that when the download request is originated from any client, the proposed framework will first attempt to evaluate the secure confidential identity of the peer nodes and once the authentication is positively accomplished, the cleaner chunk of the digital content requested by the authorized client starts downloading whereas the unauthorized clients will end-up either downloading a poisoned chunks of digital data which will render the unauthorized client to increase the download time to higher extent.

It is clearly observed in fig 9. that when the experiment is performed with 3 prominent p2p protocols e.g. BitTorrent, NapStar , and Gnutella, it can be seen that download time is reduced with every bytes of download of clean chunk of digital contents. While reverse event is witness in case of unauthorized clients as shown in fig 10, where it can be seen that the download time increase with every bits of download of poisoned chunk of data.

## VIII. CONCLUSION

The proposed system highlights a novel approach for discriminating authorized and unauthorized peer client in the network. The main intention is to discourage the attempt of any illegitimate client to download the legal and premium content of the digital file on which only the legitimate client have the rights to download. The approach discussed identifies the legal and illegal client and once successful identification is accomplished, the proposed algorithm sends the poisoned chunk of data to the unauthorized client and clean chunk of the digital data only for the genuine and legal peer node. Simulation results claims to highlights the efficiency of proposed algorithm.

## Reference

[1] http://en.wikipedia.org/wiki/File_sharing
[2] http://en.wikipedia.org/wiki/Content_delivery_network
[3] http://en.wikipedia.org/wiki/Peer-to-peer
[4] Xin Xiangjun; Xing Peixu;, Efficient Certificate-Based and Randomized Signature from Pairings, Information Engineering, 2009. ICIE '09. WASE International Conference on Issue Date: 10-11 July 2009
[5] Nicolas Christin, Andreas S. Weigend and John Chuang, Content Availability, Pollution and Poisoning in Peer-to-Peer File Sharing Networks, Proceedings of the 6th ACM conference on Electronic commerce, 2005
[6] Ruichuan Chen, Eng Keong Lua, Jon Crowcroft, Wenjia Guo, Liyong Tang, Zhong Chen, Securing Peer-to-Peer Content Sharing Service from Poisoning Attacks, Peer-to-Peer Computing , 2008. P2P '08. Eighth International Conference on Issue Date: 8-11 Sept. 2008
[7] STEPHANOS ANDROUTSELLIS-THEOTOKIS AND DIOMIDIS SPINELLIS, A Survey of Peer-to-Peer Content Distribution Technologies, ACM Computing Surveys, Vol. 36, No. 4, December 2004, pp. 335–371.

[8] Dan Boneh, Matthew Franklin, Identity-Based Encryption from the Weil Pairing, SIAM J. of Computing, Vol. 32, No. 3, pp. 586-615, 2003.
[9] Lei Guo, Songqing Chen, Shansi Ren, Xin Chen, and Song Jiang, PROP: a Scalable and Reliable P2P Assisted Proxy Streaming System, Distributed Computing Systems, 2004. Proceedings. 24th International Conference on 2004
[10] Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, A ReputationBased Approach for Choosing Reliable Resources in PeertoPeer Networks, Proceedings of the 9th ACM conference on Computer and communications security ACM New York, NY, USA ©2002
[11] D. Dumitriu,_ E. Knightly,_ A. Kuzmanovic, _ I. Stoica, and W. Zwaenepoel_,, DenialofService Resilience in PeertoPeer File Sharing Systems, Proceeding SIGMETRICS '05 Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems ACM New York, NY, USA ©2005
[12] Ton Kalker1, Dick Epema2, Pieter Hartel3, Inald Lagendijk4, Maarten van Steen5, Music2Share – Copyright-Compliant Music Sharing in P2P Systems, Proceedings of the IEEE Issue Date: June 2004
[13] Balachander Krishnamurthy, Craig Wills, Yin Zhang, On the Use and Performance of Content Distribution Networks, ACM SIGCOMM INTERNET MEASUREMENT WORKSHOP 2001
[14] Stefan Saroiu, Krishna P. Gummadi, Richard J. Dunn, Steven D. Gribble, and Henry M. Levy, An Analysis of Internet Content Delivery Systems, ACM SIGCOMM INTERNET MEASUREMENT WORKSHOP 2001
[15] Pablo Rodriguez, SeeMong Tan, Christos Gkantsidis, On the feasibility of Commercial, Legal P2P Content Distribution, ACM SIGCOMM Computer Communication Review Homepage archive Volume 36 Issue 1, January 2006
[16] Matthew Yurkewych Brian N. Levine Arnold L. Rosenberg, On the Cost Ineffectiveness of Redundancy in Commercial P2P Computing, Proceeding CCS '05 Proceedings of the 12th ACM conference on Computer and communications security ACM New York, NY, USA ©2005
[17] Kevin Walsh, Emin G¨un Sirer, Fighting PeertoPeer SPAM and Decoys with Object Reputation, Proceeding P2PECON '05 Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems ACM New York, NY, USA ©2005
[18] Runfang Zhou, Kai Hwang, GossipTrust for Fast Reputation Aggregation in Peer-to-Peer Networks, IEEE TRANSACTIONS ON KNOWLEDGEMENT AND DATA ENGINEERING (TKDE-0003-0107R1, FINALIZED FEB. 11, 2008)

# A Comparative Study on Performance Benefits of Multi-core CPUs using OpenMP

**Vijayalakshmi Saravanan[1], Mohan Radhakrishnan [2], A.S.Basavesh [2], and D.P. Kothari[3]**

**Ryerson University, Canada[1], HCL Canada, Canada[2], NITK, India[2] , VITS, Nagpur[3]**

**Abstract**

Achieving scalable parallelism from general programs was not successful to this point. To extract parallelism from programs has become the key focus of interest on multi-core CPUs. There are many techniques and programming models such as MPI, CUDA and OpenMP adopted in order to exploit more performance. But there is an urge to find the best parallel programming techniques for the benefit of performance. This article shows how the performance potential benefits the parallel programming model over sequential programming model. To support our claim, we are likely to analyze the performance in terms of execution time on both sequential and parallel implementations of naive matrix multiplication vs. Strassen's matrix multiplication algorithm using OpenMP. Our analysis results show that optimizing the code using OpenMP increases the performance than sequential implementation and outperforming well with parallel algorithms.

***Keywords:*** *Multi-core, Performance Analysis, OpenMP, Strassen's Algorithm, Parallelism.*

## 1. Introduction

In the recent years, the computer architects no longer rely on increasing single-core processor clock speed or micro architectural improvements to enhance processor performance and found it difficult in exploiting more instruction level parallelism from a single program. Thread-level parallelism could be a well-known strategy to improve processor performance. So, this results in multithreaded processors. Unfortunately, most applications are not multithreaded. Thus, adding cores results in little performance improvement. Researchers have proposed many programming languages to exploit parallelism [1] [5] [6]. These languages allowing high-level parallelism makes parallel programming easier than earlier methods.

Matrix multiplication is an important core computation in many areas of scientific computing. Normally for small multiplication we lean towards to use naive matrix multiplication algorithm which has rich data parallelism. To obtain more performance through that algorithm we parallelized them using OpenMP.

As matrix size grows the naive matrix multiplication becomes inefficient in terms of performance. For large matrices, we used Strassen's algorithm for matrix multiplication (recursive, divide and conquer approach), to enhance the performance of this algorithm on multi core architecture which has functional parallelism in its algorithm, we used OpenMP to parallelize. The results of using OpenMP in each algorithm were encouraging.

The rest of the paper as follows Section 2 describes about overview of OpenMP. Section 3. Brief about related work. Section 4 describes the algorithm and implementation methodology. Section 5 explains the tabulation of how OpenMP helps to improve performance of multi-core processors using Strassen's vs. naive algorithm. Section 6 discusses Result analysis and Discussions. Section 7 finally provides the conclusion and future work.

Hardware and Software Used:

Table 1: System specifications

| | | |
|---|---|---|
| Processor | | Intel Core i3 Dual-core |
| CPU | | 2.13GHz |
| RAM | | 4GB |
| Operating System | | Windows 7/ Ubuntu 9.04 or later |
| Soft wares | | Visual Studio 2005, GCC compiler (Linux) |

## 2. Overview of an OpenMP

The OpenMP (open multiprocessing) is an application interface platform for shared memory and it consists of set of compiler directives, library routines and environment variables that directs run time behavior. Multiprocessor programming in C or C++ on such architecture includes UNIX and windows operating systems [9]. Due to the boom in hardware speeds and the drop in hardware costs, several developers have let code optimization slip to the back of their minds. As a result, the previously developed techniques from years ago have not been updated to account for modern compiler optimization techniques or hardware features. Synthesizing a large volume of data from opposing viewpoints led to the development of a general outline to follow when optimizing code.

Many programmers will choose a language they are familiar with, even though if it's not the most effective language for the research work. Speed, flexibility, and ease of coding are a few of the major factors in deciding which language to use. The compiler will perform several optimizations faster than human programmer does. Optimization like moving constant expressions outside of loops, storing variables in registers, moving functions inline, and unrolling loops should be performed by the compiler in most cases. Parallelization of sequential programs, parallelizing compiler depend upon subscript analysis to detect data dependencies between pairs of array references inside the loop nests.

To understand the concept of OpenMP, there's a necessary to understand the concept of parallel programming. Parallel processing is done by more than one processor in parallel computing systems. Earlier multiprocessing systems always came in its own processor packaging, however recently introduced multi-core processor can contain multiple processor or cores on a single chip. In this work we achieved thread level parallelism using OpenMP and it reduces the communication cost. OpenMP is an API which acts as parallel programming model on multi-core architecture.

## 3. Related work

Prior work has studied the implementations of Strassen's matrix multiplication algorithms in many programming languages such as C, C++ and Java [4]. But there is a need to understand the parallel programming and its implementation methodologies on multi processors system in order to improve the performance. OpenMP is the well-known parallel programming techniques for multiprocessing environment [2]. There are varied hardware and software techniques adopted for performance enhancements.

One of the traditional methods to achieve more performance is to increase the clock frequency. There are different kinds of heterogeneous pipeline models have been discussed by many researchers. Latch based pipelines are most commonly used pipelines in asynchronous circuit pipeline models [10]. Fine grained and coarse grained pipeline structure which focuses on cell gate implementation were introduced and improved by many computer architecture researchers [7] [3].

Recently, SR (Self Resetting) latches were proposed by [8] which resolve the power consumption problem and reduces the data path power consumption. Kunkel and Smith studied the performance improvement using gate level logic circuits. As there's tremendous improvement in silicon technology the problems of clocking, range of transistors on chip, and will increase the complexity on chip. Therefore there's an urge to find the software or hardware algorithm to solve this issue.

## 4. Algorithm and Implementation Methodology

### (A) Sequential naive algorithm

We used both sequential and parallel versions of naive and Strassen's algorithm to analyze the performance shown in Figure.6. The pseudo code for the naive algorithm of matrix multiplication of matrix a (n*n) and matrix b (n*n) to give a matrix c (n*n) is shown in Figure.1 [14].

### (B) Parallel naive algorithm using OpenMP

The pseudo code for the naive algorithm of matrix multiplication of matrix a (n*n) and matrix b (n*n) to give a matrix c (n*n) is shown in Figure.2 [14].
It can be viewed as divide and conquer method algorithm, suppose we wish to compute the product of

$$C = A * B \text{ ------ } (1)$$

```
SQUARE-MATRIX-MULTIPLY (A, B, C)
N = A.rows
For I = 1 to N
For j = 1 to N
C_ij = 0
For k = 1 to N
C_ij = C_ij + A_ik * B_kj
Return C
```

Fig. 1: Pseudo code for sequential naive algorithm

```
SQUARE-MATRIX-MULTIPLY (A, B, C)
#pragma omp parallel for default (none) shared (A, B, C, N) private (I, j)
N = A.rows
For I = 1 to N
For j = 1 to N
C_ij = 0
For k = 1 to N
C_ij = C_ij + A_ik * B_kj
Return C
```

Fig. 2: Pseudo code for parallel naive algorithm

In Equation [1], where each of A, B and C are n*n matrices. Assuming that n is an exact power of two, we tend to divide each of A, B and C into four n/2*n/2 matrices, which can be written as shown in Figure.3.

| | | | | | |
|---|---|---|---|---|---|
| | a | b | | e | f |
| A= | c | d | B= | g | h |
| | r | s | | ae+bf | ag+bh |
| C= | t | u | C= | ce+df | cg+dh |

Fig. 3: Strassen's serial algorithm

## 4.1 Strassen's Serial Algorithm

As we can see from the above, the serial algorithm has recursion in the algorithm. We can hardly see the data parallelism except in adding and subtracting sub matrices. If the matrix size is greater than the threshold value multiply them recursively, if not use the traditional matrix multiplication algorithm.

Construct C using the intermediate matrices. But, if we look more as shown in Figure.4 and we get to understand that P1….P7 goes on recursively and independently thus we get functional parallelism.

## 4.2 Strassen's Parallel Algorithm

Initially, we implemented our program through task pool model to compute P1; P2...P7 and an independent multiplication task can be executed in parallel with N jobs at a time. For example, when 49 jobs are running with N cores machine (where $N= 2^N$) there is a chance to execute 48, and would run simultaneously with 1 job would be left later execution. Thus, it leads to more processor utilization. So as to avoid this issue, it's better to split the last task further as shown in Figure 7.

In OpenMP the sections construct is the easiest way to get the different threads to carry out different kinds of work. Since, it permits us to specify many different code regions and each of which will be executed by one of the threads in OpenMP with Strassen's matrix multiplication are shown in Figures 4 and 5 [12] [13].

Fig.4 STRASSEN'S MATRIX MULTIPLICATION

$$A = \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{21} \end{matrix}$$

$$B = \begin{matrix} B_{11} & B_{12} \\ B_{21} & B_{21} \end{matrix}$$

$$C = \begin{matrix} C_{11} & C_{12} \\ C_{21} & C_{21} \end{matrix}$$

$$C = \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} * \begin{matrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{matrix}$$

$C_{11} = A_{11}*B_{11} + A_{12}*B_{21}$

$C_{12} = A_{11}*B_{12} + A_{12}*B_{22}$

$C_{21} = A_{21}*B_{11} + A_{22}*B_{21}$

$C_{22} = A_{21}*B_{12} + A_{22}*B_{22}$

Fig.5 Strassen's Parallel Algorithm

Evaluate the intermediate matrices:

$P_1 = (A_{11} + A_{22}) (B_{11} + B_{22})$

$P_2 = (A_{21} + A_{22}) B_{11}$

$P_3 = A_{11} (B_{12} – B_{22})$

$P_4 = A_{22} (B_{21} – B_{11})$

$P_5 = (A_{11} + A_{12}) B_{22}$

$P_6 = (A_{21} – A_{11}) (B_{11} + B_{12})$

$P_7 = (A_{12} – A_{22}) (B_{21} + B_{22})$

Construct C using the intermediate matrices:

$C_{11} = P_1 + P_4 – P_5 + P_7$

$C_{12} = P_3 + P_5$

$C_{21} = P_2 + P_4$

$C_{22} = P_1 – P_2 + P_3 + P_6$

## 5. Tables for Performance Analysis of Strassen's vs. Naïve Algorithm

Table 2.Tabulation of performance analysis on Strassen's vs. Naïve algorithm

| Matrix Size (n) | Naive Serial | Naive Parallel | Strassen's Serial | Strassen's Parallel |
|---|---|---|---|---|
| 500 | 1.23 | 0.88 | 2.4 | 2.2 |
| 1000 | 13.2 | 8.11 | 6.2 | 5.4 |
| 1500 | 59.35 | 31.92 | 12.6 | 7.8 |
| 2000 | 99.62 | 79 | 22.3 | 10.81 |
| 2500 | 279.23 | 178.5 | 40.23 | 19.46 |
| 3000 | 394.5 | 316.62 | 62.35 | 28.86 |

In our work, we have tested the sequential version and parallel version (using OpenMP) for both naive and Strassen's algorithm for matrix multiplication. The execution time was taken using OpenMP run time library function omp_get_wtime () which gives time in seconds with double precision. Using OpenMP the parallelism can be achieved through the evaluations of intermediate matrices P1, P2 ... P7 which are independent as shown in Figure.5 and hence, it will be computed in parallel through Strassen's parallel matrix multiplication. Comparing the serial and parallel version of naive algorithm we got significant results from matrix sizes of more than 100*100 due to time consumed in thread synchronizing for smaller matrices. For Strassen's algorithm initially the performance was quite disappointing for smaller matrices and as the matrix sizes became larger than 500*500 the performance improved compared to naive multiplication and also the parallel version of Strassen's algorithm and the execution time is shown in Table 2 (Time in sec). The results graphs are depicted in Fig.8 and Fig.9.

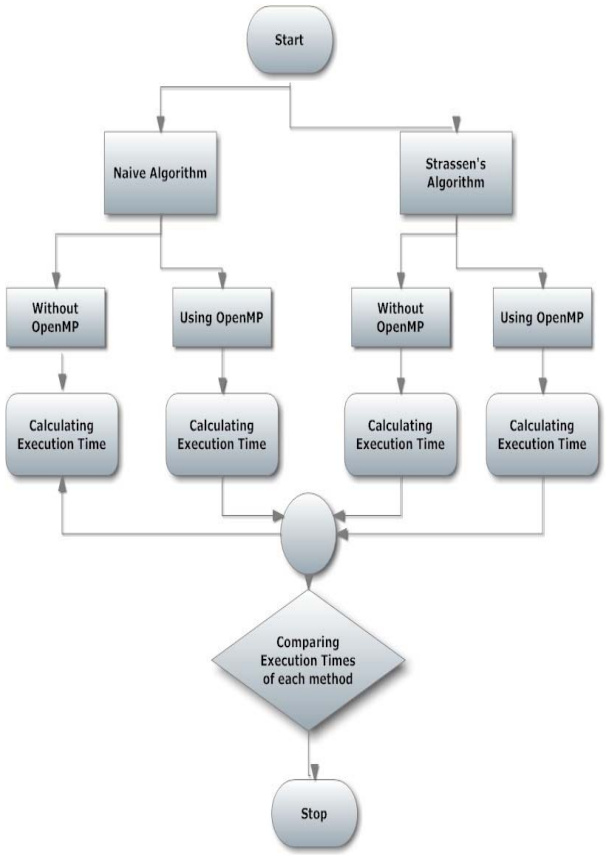



Fig. 6: Schematic flow diagram

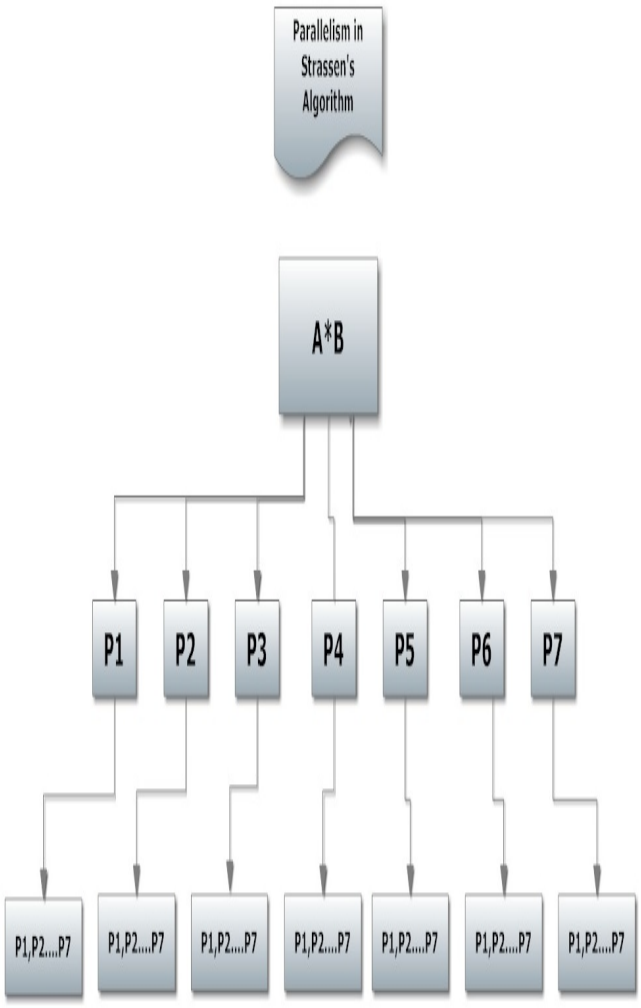



Fig. 7: Flow diagram of Strassen's algorithm

### Results Graphs

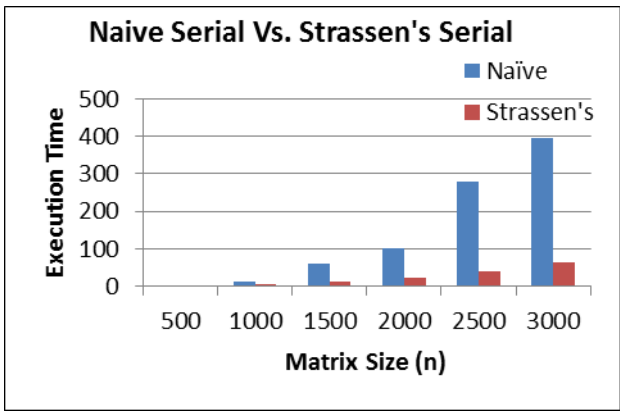



Fig.8 Naive vs. Strassen's Serial Algorithm

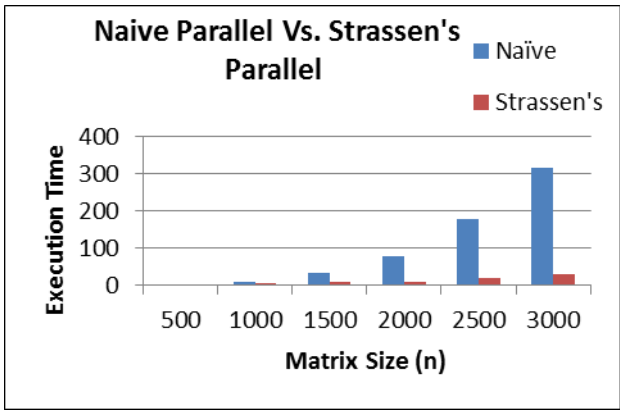



Fig.9 Naïve vs. Strassen's Parallel Algorithm

## 6. Result analysis and Conclusion

Based on our study, we have presented the execution time of both serial and parallel execution of naive and Strassen's algorithm for matrix multiplication. We arrive at the following conclusions:

(a) we see that parallelizing the serial algorithm using OpenMP has increased the performance a lot (b) For CPU cores OpenMP provides a lot of performance increase and parallelization can be done with minimal changes and, (c) we tend to observe that though Strassen's algorithm (both parallel and serial) definitely consumes a lot more memory than serial algorithm, but the performance is much better than the traditional matrix multiplication algorithm due to its reduced operations (d) overall we conclude that for large matrices we can apply Strassen's algorithm and for smaller matrices we must apply naive algorithm. And using OpenMP for both algorithms we achieved much better performance than serial implementation.

## 7. Future Enhancements

Due to time constraints, this work has been carried out on dual-core machine with matrix multiplication alone, but it can be extended by using a variety of matrix types - dense, sparse, large data, complex numbers, etc. to characterize our comparison to understand the better performance benefits of the OpenMP techniques. Besides there's a scope to look at the energy consumption of assorted algorithms and its impact on performance enhancement.

**Acknowledgments**

The authors would like to convey immense thankfulness to IASc (Indian Academy of Sciences, Bangalore) and all the anonymous reviewers for their valuable comments.

## References

[1] Brian D et al. Carlstrom.”The Atomos Transactional Programming Language”. In ACMSIGPLAN2006 Conference on Programming Language Design and Implementation. June 2006.
[2] Barbara Chapman. ”Managing Multi-core with OpenMP (Extended Abstract)”. In Proceedings of the 15th European PVM/MPI Users’ Group Meeting on “Recent Advances in Parallel Virtual Machine and Message Passing Interface”, pages 3–4, Berlin, Heidelberg, 2008. Springer Verlag.
[3] Zenil Chavez. ”Applied Parallel Computing”. IEEE Distributed Systems Online, 5, 2004.
[4] Thomas H.et al. Cormen. ”Introduction to Algorithms”. McGraw-Hill Higher Education, 2nd edition, 2001.
[5] Matteo et al. Frigo. ”The implementation of the Cilk 5 multithreaded language”. In Proceedings of the ACMSIGPLAN1998 conference on Programming language design and implementation, PLDI ’98, pages 212–223, New York, NY, USA, 1998. ACM.
[6] Michael I. et al. Gordon. ”A stream compiler for communication exposed architectures”. SIGARCH Comput.Archit. News, 30:291–303, October 2002.
[7] Shi Jung Kao. ”Managing C++ OpenMP code and its exception handling”. In Proceedings of the OpenMP applications and tools 2003 international conference on OpenMP shared memory parallel programming, WOMPAT’03, pages 227–243, Berlin, Heidelberg, 2003. Springer-Verlag.
[8] Quin Michael J.”Parallel programming in C with MPI and OpenMP”. McGraw Hill Inc., 2004.
[9] Venkatesan Packirisamy, Harish Barathvajasankar, S Sarholz in Proceedings of the 3rd international workshop on OpenMP "A Practical Programming Model for the Multi-core Era" (2008).
[10] Alex Vrenios. ”A Tutorial on Parallel Systems Development”. IEEE Distributed Systems Online, 5, 2004.
[11] www.OpenMP.org.
[12]http://ace.cs.ohiou.edu/~razvan/courses/cs404/lecture12.pdf.
[13] Steven Huss-lederman, Elaine M. Jacobson, J. R. Johnson, Anna Tsao, Thomas Turnbull “Strassen’s Algorithm for Matrix Multiplication: Modeling, Analysis, and Implementation” In Proceedings of Supercomputing '96.
[14] John Burckardt, Paul Puglielli Pittsburgh Supercomputing Center, “MATMUL: An Interactive Matrix Multiplication Benchmark “.

**Vijayalakshmi Saravanan** is an Assistant Professor (Sr); VIT University, India .She is a recipient of Erasmus Mundus (EURECA) Programme as an Exchange student from India at Malardalen University, Sweden. She holds a Bachelor of Engineering Degree in Electrical and Electronics Engineering and Master of Science Degree in Information Technology from Bharathiar University & Manonmaniam Sundaranar University (Now Anna University), India. Currently, she holds a position as visiting researcher at Ryerson University, Canada. Her research interests include Multi-core Low Power Design Exploration, Power-Aware Processor Design, and Computer Architecture. She has taken part of her research studies one course work at University of Rochester, USA. She is serving as a Technical Evangelist for Asia Open Source Software Community, CICC, and Japan and all over Asian Countries. She is a Member of IEEE, ACM, CSI and a Board member of N2WOMEN (Networking Networking Women) IEEE/ACM Women in Engineer and she is a Chair for IEEE-WIE VIT affinity group, India. She can be reached at viji@ieee.org.

**Mohan Radhakrishnan** is currently working as a Sr.Technical Architect in HCL Canada. He has more than ten years of technical experience in designing, administrating and supporting Microsoft enterprise and VMware environments. He is currently working on R&D level projects in data center server and network implementation, support and administration, thorough grasp of development principles and best practices. He is also a Member of IEEE and VMware.

**Dr. D.P. KOTHARI** is a Senior Professor and Advisor to the Chancellor, VIT University, Vellore and named IEEE fellow in 2011. Earlier, he was Head, Centre for Energy Studies, IIT Delhi (1995-97), and Principal, Visvesvaraya Regional Engineering College, Nagpur (1997-98). He has been Director i/c, IIT Delhi (2005) and Deputy Director (Administration) (2003-06). Earlier, (1982-83 and 1989), he was a visiting fellow at RMIT, Melbourne, Australia. He obtained BE, ME and PhD

degrees from BITS, Pilani. He is a Fellow of the Institution of Engineers (India), Fellow of National Academy of Engineering (FNAE), Fellow of National Academy of Sciences (FNASc), Life Member ISTE (LMISTE). Professor Kothari has published/presented 640 papers in national and international journals/conferences. He has authored/co-authored 22 books including Power System Optimization, Modern Power System Analysis, Electric Machines, Power System Transients, Theory and Problems of Electric Machines, Renewable Energy Sources and Emerging Technologies, and Power System Engineering. His research interests include Optimal Hydro-thermal Scheduling, Unit Commitment, Maintenance Scheduling, Energy Conservation (loss minimization and voltage control), and Power Quality and Energy Systems Planning and Modeling. He has received the National Khosla award for Lifetime Achievements in Engineering for 2005 from IIT Roorkee. The University Grants Commission (UGC) has bestowed UGC National Swami Pranavananda Saraswati award for 2005 on Education for outstanding scholarly contribution. The World management congress, New Delhi conferred Life time achievement award for "Educational Planning and Administration" on 30th December 2009.

# Improving Internet Quality of Service through Active Queue Management in Routers

**Gamal Attiya[1] and Heba El-Khobby [2]**

**[1] Dept. of Computer Science and Engineering,
Faculty of Electronic Engineering,
Minoufiya University, Egypt**

**[2] Dept. of Electronics and Electrical Communications Engineering,
Faculty of Engineering, Tanta University, Egypt**

**Abstract**

The traffic characteristics of real-time and non real-time applications require a certain Quality of Service (QoS) from the Internet in terms of bandwidth, delay, packet loss, fairness and jitter. However, most of the current Active Queue Management (AQM) algorithms at the internet routers do not guarantee QoS for real time traffics such as video and audio. This is because; most of the algorithms handle different packets of different traffics by the same strategy. In this paper, we propose a new AQM strategy to guarantee QoS for real time traffics. The proposed strategy uses three queues at the internet routers, each of which handles a single class of traffic. Where, the arriving packets are queued according to their class type. Additionally, the queued packets are scheduled according to a predefined weight. The proposed algorithm is evaluated and compared with the most recent algorithms by using the Network simulator NS-2.

***Keywords:*** *Active Queue Management, Packet Scheduling, QoS, Multimedia, Congestion Control*

## 1. Introduction

In the past few years, the Internet is moving from traditional data communication network for transferring text-based messages such as file transfer (FTP), email and web browsing, to an underlying communication network for multimedia applications such as IP telephony, interactive video conferencing, video on demand and online games. However, with the increasing volume of traffic from both traditional and multimedia applications, a serious problem, called congestion collapse, arises [1]. This problem leads to performance degradation particularly for real-time applications.

Over years, continuous efforts are being done and several congestion control mechanisms are being developed to avoid this problem. The congestion control methodologies can be categorized into End-to-End protocols and Active Queue Management (AQM) algorithms. The End-to-End congestion control protocols are implemented at the end systems based on the Transmission Control Protocol (TCP) [2]. The end system's TCP protocol adjusts its sending rate by adjusting its window size using Additive Increase and Multiplicative Decrease (AIMD) method according to the signaled congestion (e.g. packet loss). Several algorithms to the End-to-End congestion control are investigated and evaluated in [3].

On the other hand, the Active Queue Management (AQM) algorithms are implemented at the internet routers, wherein a router signals impending congestion by dropping packets. Queues are used in routers to absorb transient bursts in incoming packet rates, allowing the router sufficient time for packet transmission. When the incoming packet rate is consistently higher than the router's outgoing packet rate, the queue size will increase, eventually exceeding available buffer space. When the buffer is full, some packets must be dropped. The most well-known active queue management mechanism is Random Early Detection (RED) [4]. RED operates based on an average queue length that is calculated using an exponential weighted average of the instantaneous queue length. It uses the weighted-average queue size and thresholds to detect impending congestion, and randomly drops incoming packets as the average queue size exceeds a minimum threshold. RED drops packets with a probability depending on the average length of the queue. The drop probability increases from 0 to a maximum drop probability as the average queue size increases from a minimum threshold to the maximum threshold. If the average queue size goes above the maximum threshold, all packets are dropped.

The RED was developed to keep the average queue size small, increase throughput, reduce burst loss and global synchronization. However, it is known that RED's effectiveness is heavily dependent on the setting of its control parameters. Moreover, the average queue length of

RED in steady state depends on the number of active TCP connections. Some of the difficulties in RED configuration are explained in [5]. Several variants of RED have been proposed to address the performance problems of RED. These variants are; Fair Random Early Drop (FRED) [6], Stabilized RED (SRED) [7], Gentle RED (GRED) [8], DS-RED [9], CHOCKe [10], Dynamic-RED (DRED) [11], Adaptive RED [12], MRED [13], Random Early Mark (REM) [14], Green [15], Blue [16], PD-RED [17], LRED [18], HRED [19], ARED [20] and AutoRED [21] and Adaptive CHOKe [22]. Classification and performance evaluation of the different variants of the AQM schemes are presented in [23].

Although many variants were developed, the current Internet only provides best-effort service for non-real-time traffics. But, it does not guarantee quality of service (QoS) for multimedia or real-time traffics [24-26]. This paper presents a new strategy to the RED AQM in order to guarantee quality of service (QoS) for the real-time traffics. The approach uses three queues each of which handles a single class of traffic: video, audio or data. The queued packets are then scheduled according to a predefined static weight. The proposed strategy is investigated and compared with the RED AQM by using the network simulator NS-2 [27-29].

The rest of this paper is organized as follows. Section 2 describes the problem in some details. A description of the proposed algorithm is presented in Section 3. Section 4 presents the simulation result that showing the effect of the proposed approach on the throughput, delay and packet loss of different traffics. Section 5 summarizes our conclusions and future work.

## 2. Problem Definition

Multimedia applications have different performance constraints than do traditional applications. Traditional applications are very sensitive to lost packets. Multimedia applications, on the other hand, can tolerate some data loss, but are very sensitive to variance in packet delivery, called jitter. In the absence of jitter and packet loss, video frames can be played as they are received, resulting in a smooth play-out. However, in the presence of jitter, the user would see the frozen image of the most recently delivered frame until the tardy frame arrived. The tardy frame would then be played only briefly in order to preserve the timing for the subsequent frame.

The Active Queue Management (AQM) algorithms control packet loss by Early Congestion Notification (ECN) marks to signal congestion. It requires that the end-hosts recognize dropped packets and respond by retransmitting the dropped packets and reducing their rate of transmission. In the Internet, TCP recognizes packet loss as an indicator of network congestion, and reduces transmission rate. Hence, to date, active queue management appears promising since by far the predominant transport protocol on the Internet is TCP.

Traditional applications use TCP to guarantee that lost packets are retransmitted. Unfortunately, detecting and retransmitting lost packets causes considerable jitter, making TCP unattractive to multimedia applications. So, most non-TCP flows (real time traffics) use User Datagram Protocol (UDP) without employing end-to-end flow and congestion control. But, UDP get an unfair share of network bandwidth when there is congestion. This unfairness occurs because many non-TCP flows do not reduce transmission rates while the TCP flows are forced to transmit data at their minimum rates. Even worse, typical active queue management policies apply the same drop rate to each flow. Thus, when UDP coexists with TCP, it induces not only a congestion collapse problem but also an unfairness problem that each flow cannot get the same treatment, causing an unstable Internet and lower link utilization. In this paper, we propose an approach to the RED-AQM to give equal importance to both the throughput and delay requirements of an application.

## 3. Proposed Algorithm

The main idea of the proposed algorithm is illustrated in Figure 1. As shown in the figure, three queues are used in the Internet router to queue packets that arrive from different sources, where each queue is assigned a single type of traffic. Packets from different sources are first arrived at a classifier which discriminates between different types of traffics. The classifier has to distinguish packets belonging to a traffic type via a certain field in the packet header. As shown in Figure 1, video traffic is classified as high priority flow and inserted into the first queue, audio traffic is classified as medium priority flow and inserted into the second queue, and other flows is classified as low priority flows and inserted into the third queue.

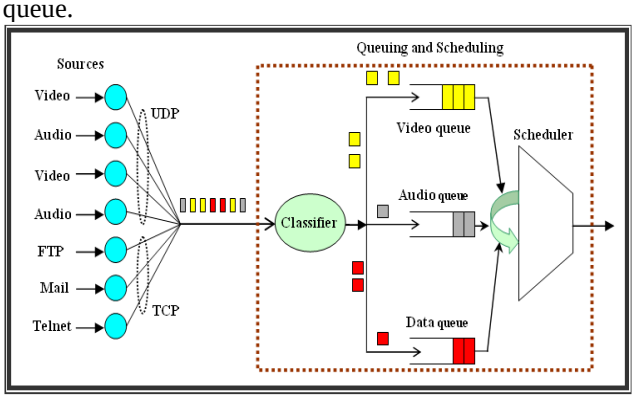



Figure 1: Illustration of Proposed Algorithm

The problem now is how to serve different packets that are queued in the different queues. The simplest solution is the priority based scheduling, where packets in different queues are served with different priorities. In other words, packets in higher priority queues are served first and other packets from a certain queue are scheduled only if the higher priority queues are empty. In addition, with each priority queue, packets are scheduled in FIFO strategy. This method would significantly reduce the delay of the high priority packets (delay sensitive or real-time traffics such as video and audio). However, if the amount of high priority traffic is excessive, the low priority queue may not get any service until the high priority traffic is completely served. In this case, the queues allocated to lower priority traffic (not delay sensitive or non real-time traffics such as FTP and Mail) may overflow. As a result, the low priority traffic may experience a large delay or, in the worst case, a complete resource starvation.

To prevent resource starvation, the service rate of the high priority traffic should be limited. In the proposed approach, each type of traffic is queued in a separate queue and the ratio of service for each type is determined. In other words, a weight is assigned to each queue and packets are scheduled according to the predefined weight. For example, the service ratio may be 3:2:1; this means that three packets are served from high priority video traffic, two packets are served from medium priority audio traffic and one packet is served from low priority traffic (traditional data). The main advantage of this strategy is to allow TCP sources to utilize the available bandwidth efficiently after the high priority traffics consume their available capacity. Indeed, critical time applications overcome the end-to-end latency because they can be served with higher priority. In the proposed approach the service ratio is 1:1:1.

## 4. Simulation Results

The proposed strategy is coded in C++ and incorporated into the Network Simulator NS-2 to be used as an AQM scheme in the router. Additionally, the performance of the proposed algorithm is evaluated and compared with the RED AQM for handling real time video and audio traffics.

### 4.1. Simulation Environment

Our design and implementation are carried out in NS-2 [27-29]. It is a popular Wide Area Network simulator developed at the University of California, Berkeley but used by many others for a wide variety of network research. NS supports most of the common IP network components, including TCP (Tahoe, Reno and Vegas) and UDP transport agents, and several queue management mechanisms, including RED. Unfortunately, NS-2 considerably lacks support for multimedia applications, only providing a basic mechanism to build Constant Bit Rate (CBR) media streams. NS does not support streaming Variable Bit Rate (VBR) multimedia, such as an MPEG or Real Video. VBR applications are required for responsive multimedia applications that must maintain their strict timing constraints. Thus, a further contribution of this work is the NS-compatible source code for two flow multimedia applications; video and audio traffics.

Figure 2 shows the network topology and the application flows that used for the simulation. The topology represents a simple network bottleneck configuration. It has two multimedia (MM) sources represent real time video and audio traffics, and six FTP sources represent non real time traffics. The network links are labeled with their bandwidth capacity and propagation delay. Each source as well as each destination is connected through a link has capacity 10 Mb/s and propagation delay 5 ms. The bottleneck link has 6 Mb/s bandwidth and 20 ms propagation delay. For transferring data, the FTP uses TCP connection while the video and audio traffics use UDP connection as the underlying transport agent. All the TCP agents were set to have a congestion window size of 20 packets and a packet size of 1500 bytes. The UDP agents also have a packet size of 1500 bytes and a rate of 1.5 Mbps for video traffic and 384 kbps for audio traffic.

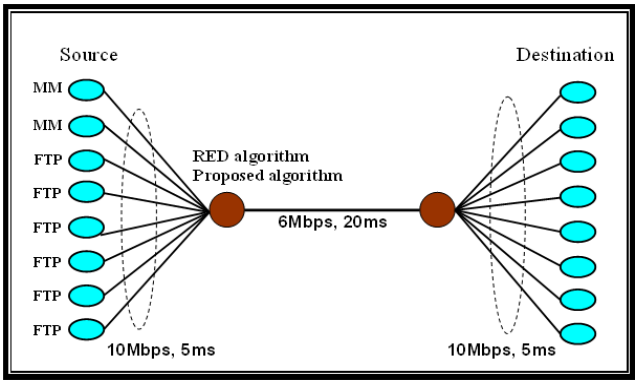
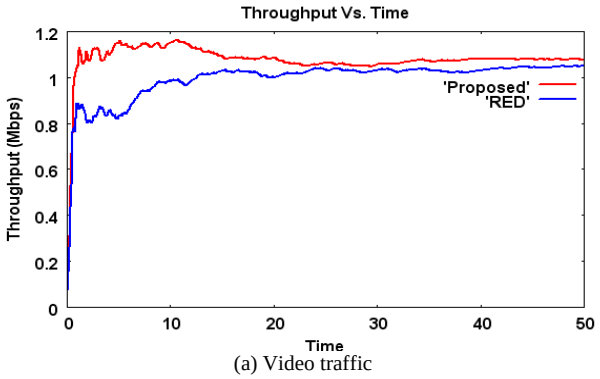



Figure 2: Network Topology and Flows

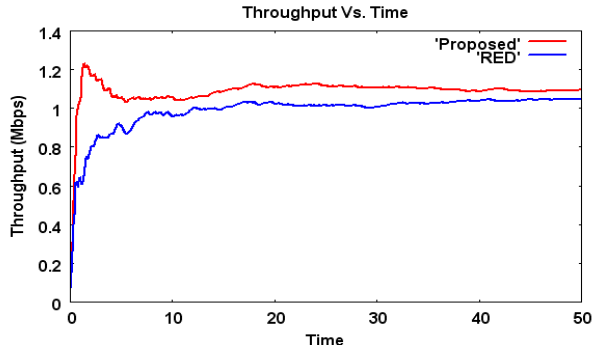
In the following, we study the effect of the proposed algorithm and the RED-AQM on the throughput, packet loss and the average delay. All measurements were occurred at the router, considering the simulation time 50 sec. The router uses the RED AQM with 20 packets long queue. The parameter settings of the RED are that the minimum threshold = 5 packets, maximum threshold = 15 packets, average queue weight = 0.002 and the probability $max_p$ =0.1.

### 4.2. Effect of the Algorithms on Throughput

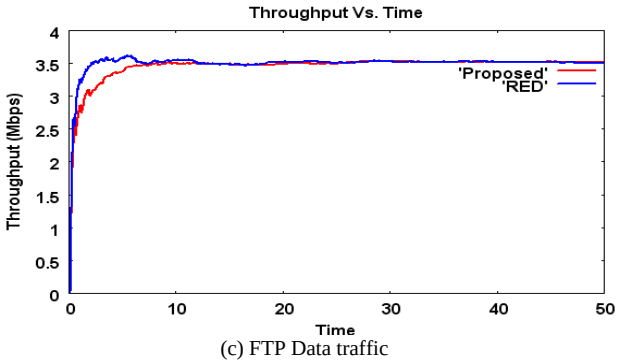
Figure 3 shows the throughput for each type of traffic when using the RED and the proposed algorithm with service ratio 1:1:1. From the figure, it is shown that, the proposed algorithm (*denoted by Proposed*) provides higher throughput than the RED algorithm (*denoted by RED*) for both video and audio traffics. For FTP flows, the throughput when using the proposed algorithm is identical to that of using the RED algorithm. This indicates that the proposed approach improves the throughput of real time traffics without affecting on the non real time traffics.




(a) Video traffic




(b) Audio traffic

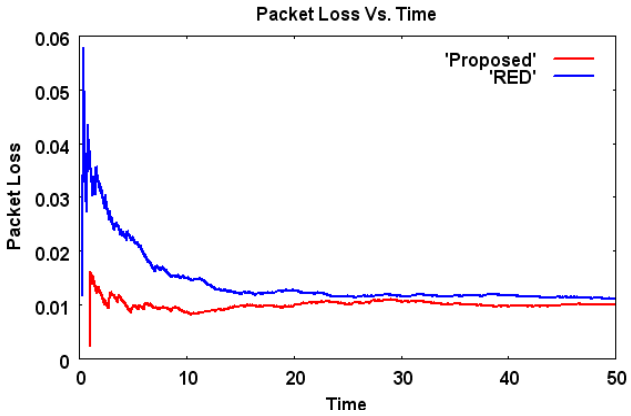



(c) FTP Data traffic

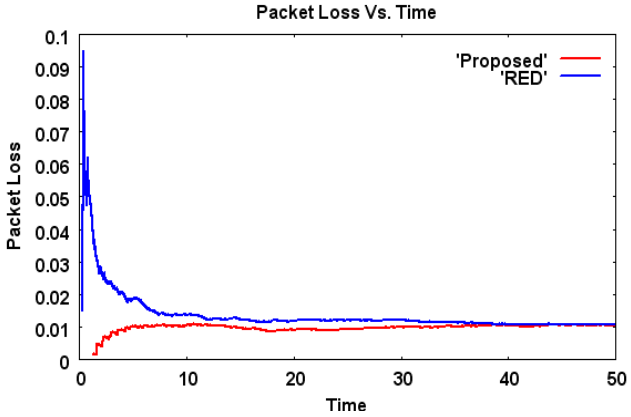
Figure 3: Throughput vs. time for the RED and the proposed algorithm.

### 4.3. Effect of the Algorithms on Packet Loss

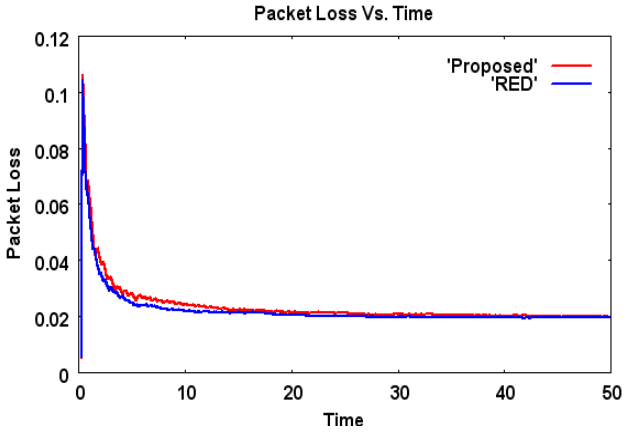
Figure 4 shows the packet loss ratio for each type of traffic when using the RED and the proposed algorithm with service ratio 1:1:1.




(a) Video traffic




(b) Audio traffic



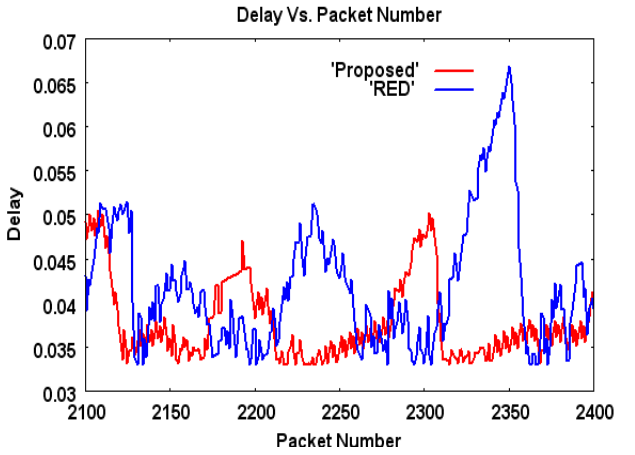

(c) FTP data traffic

Figure 4: Packet Loss vs. time for the RED and the proposed algorithm.

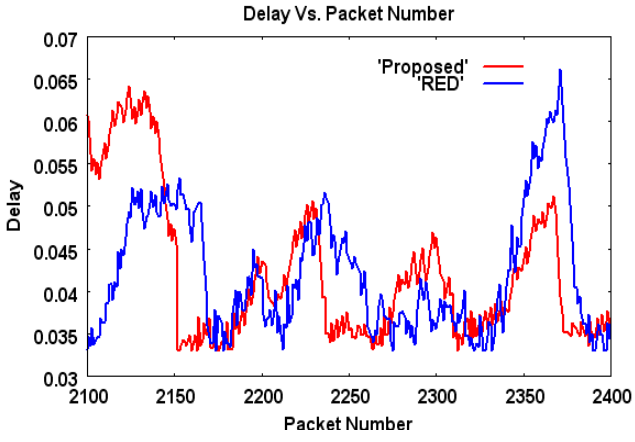
From the figure, it is shown that, the packet loss ratio when using the proposed algorithm (*denoted as Proposed*) is less than the ratio of the packet loss of the RED algorithm (*denoted as RED*) for both the real time video and audio traffics. For FTP flows, the packet loss ratio remains identical to the RED algorithm. This indicates that the proposed approach decrease the packet loss of real time traffics without affecting on the non real time traffics.

## 4.4. Effect of the Algorithms on Delay

Figure 5 shows the packet delay experienced by each type of traffic when using the RED and the proposed algorithm with service ratio 1:1:1. From the figure, it is shown that, the average delay experienced by the proposed algorithm (*denoted as Proposed*) for both Video and Audio traffics is less than that of the RED algorithm (*denoted as RED*).




(a) Video traffic




(b) Audio traffic

Figure 5: Delay vs. Packet Number for the RED and the proposed algorithm.

## 4.5. Discussion

An important question arises now, why the service ratio 1:1:1 improves performance although the behavior of the router with this ratio seems like using one queue with scheduling strategy FIFO?

The service ratio 1:1:1 improves performance because using one queue with FIFO scheduling strategy does not guarantee fairness between different traffics. Since traditional data such as FTP arrives at the internet router in burst, then many packets of the same type will be queued in the buffer. Thus, during service, one type of traffic will be served for long period of time based on the size of the burst. But, queuing packets into three different queues and using the scheduling strategy 1:1:1 give the chance of service to the different types of traffics, resulting in fairness between different traffics.

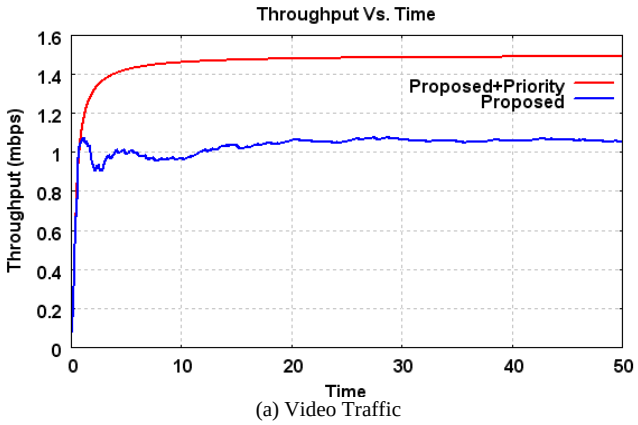
Additional improvement to the real time traffics may be achieved by serving different packets of different traffics by different weights. The problem is how to specify this weight. An approach is to predefine the weight statically with different priorities. For example, the service ratio may be 3:2:1. Another method is to define the weight dynamically based on the average queue size. That is, starting with the service ratio 1:1:1 and during each cycle test the average size of the queues and then calculate the new service ratio for the next cycle based on the current average queue size.

## 4.6 Effect of Static Priority Based Scheduling

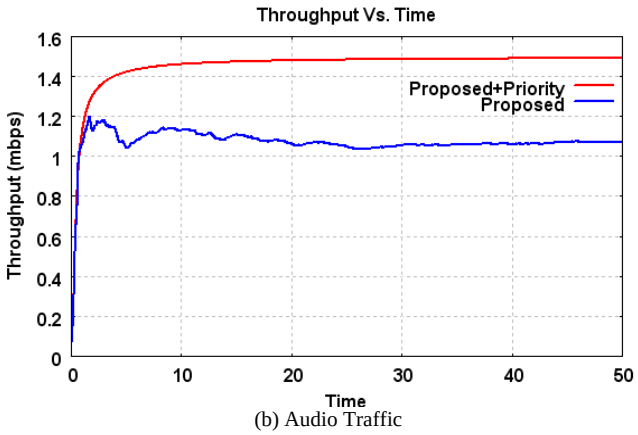
In this section, we study the effect of assigning different priorities to different types of packets. The purpose is to compare the static priority based scheme with non priority based scheme and to study the effect of giving priority to interactive traffics over the ftp traffic. The service ratio is considered as 3:2:1, i.e., for every three packets of video, two audio packets and one data packet will be served.

Figure 6 shows the throughput for different traffics using the proposed algorithm with priority (*denoted by Proposed+Priority*) and without priority (*denoted by Proposed*). The figure shows that the proposed algorithm with priority provides higher throughput than the proposed algorithm without priority for the multimedia flows but the throughput decreases for FTP flow.

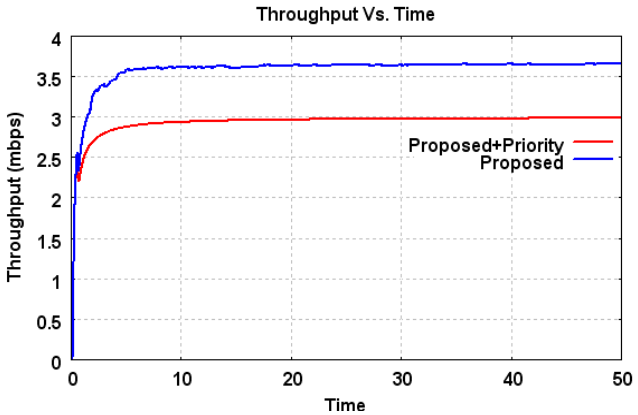
Figure 7 shows the delay of different traffics using the proposed algorithm with priority (denoted by *Proposed+Priority*) and without priority (denoted by *Proposed*). The figure shows a significant reduction in delay for both Video and Audio traffics when using the proposed algorithm with priority.



(a) Video Traffic

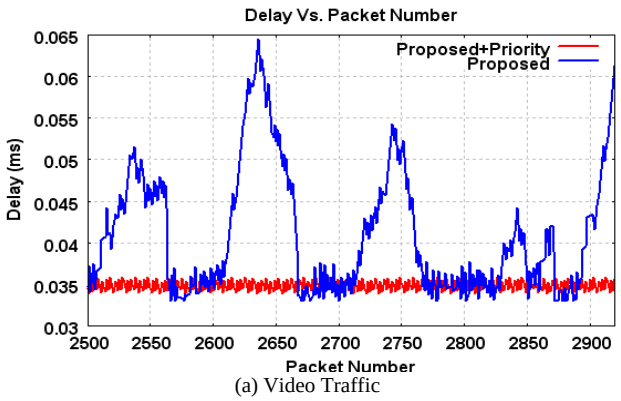


(b) Audio Traffic

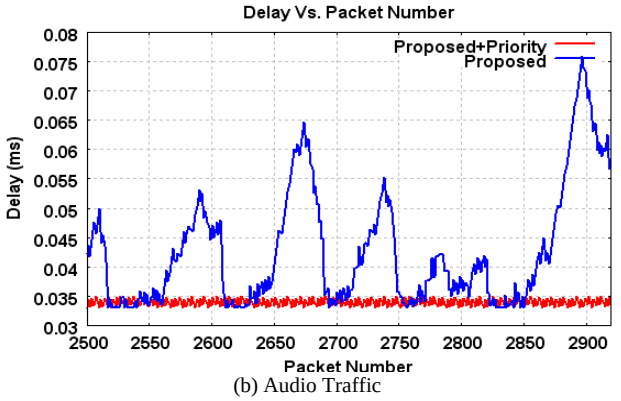


(c) FTP data Traffic

Figure 6: Throughput vs. time for the proposed algorithm with/without priority.



(a) Video Traffic



(b) Audio Traffic

Figure 7: Delay vs. Packet Number for the proposed algorithm with/without priority.

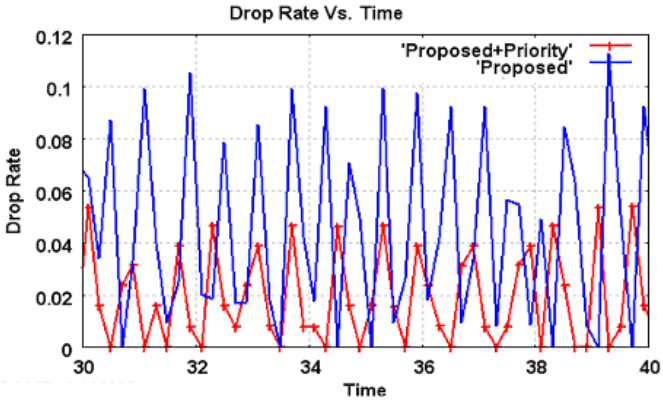



Figure 8: Packet drop rate vs. time for the proposed algorithm with/without priority.

Figure 8 shows the packet drop rate when using the proposed algorithm with priority (denoted by *Proposed+Priority*) and without priority (denoted by *Proposed*). The figure shows that the proposed algorithm with priority decreases the drop rate comparing to the proposed algorithm without priority.

## 5. Conclusions and Future Work

In this paper, a queuing discipline is proposed to the active queue management to deal with both the best-effort flows

and multimedia flows. The proposed algorithm uses three priority queues each of which handles a single class of traffic. Additionally, the queued packets are scheduled according to predefined weights. The proposed approach prevents starvation of lower priority traffic (i.e., best effort traffic) while satisfying the QoS requirements for higher priority traffics. This would result in a significant decrease in the delay and loss of the real time packets. Our future work will be concerned on the dynamic weight adjustment of packet scheduling to achieve more improvement to the real time traffics.

## References

[1] V. Jacobson, and M. J. Karels, "Congestion Avoidance and Control," Proceedings of ACM SIGCOMM, Vol.18, No. 4, pp. 314-329, August 1988.

[2] S. Floyd, and K. Fall, "Promoting the use of End-to-End Congestion Control in the Internet," IEEE/ACM Transactions on Networking, Vol.7, (4), pp. 458-472, August 1999.

[3] Hanaa A. Torkey, Gamal M. Attiya, and I. Z. Morsi, "Performance Evaluation of End-to-End Congestion Control Protocols," Minufiya Journal of Electronic Engineering Research (MJEER), Vol. 18, No. 2, July 2008.

[4] Sally Floyd and Van Jacobson, "Random Early Detection Gateways for Congestion Avoidance," IEEE/ACM Transactions on Networking, Vol.1, pp. 397-413, August 1993.

[5] V. Firoiu and M. Borden, "A Study of Active Queue Management for Congestion Control," Proceedings of IEEE INFOCM 2000, Vol. 3, pp. 1435-1444, March 2000.

[6] D. Lin and R. Morris, "Dynamics of Random Early Detection," Proceedings of ACM SIGCOMM Conference, Vol. 27, No. 4, pp. 127-137, 1997.

[7] T. J. Ott, T. V. Lakshman, and L. Wong, "SRED: Stabilized RED," Proceedings of IEEE INFOCOM '99, pp. 1346–1355, Mar. 1999.

[8] S. Floyd, "Recommendations on using the gentle variant of RED," May 2000, available at http://www.aciri.org/floyd/red/gentle.html.

[9] B. Zheng, and Mohammed Atiquzzaman, "DSRED: An Active Queue Management Scheme for Next Generation Networks" Proceedings of 25th IEEE conference on Local Computer Networks (LCN) 2000, November 2000

[10] R. Pan., B.Prabhakar, and k.Psounix, "CHOKe, a Stateless Active Queue Management Scheme for Approximating Fair Bandwidth Allocation", IEEE INFOCOMM, Feb 2000.

[11] J. Aweya, M. Ouellette, and D. Y. Montuno, "A control theoretic approach to active queue management," Computer Networks, Vol. 36, pp. 203–235, 2001.

[12] Sally Floyd, Ramakrishna Gummadi, and Scott Shenker, "Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management," Under submission, http://www.icir.org/floyd/papers/adaptiveRed.pdf, 2001.

[13] J. Koo., Byunghun Song., Kwangsue Chung., Hyukjoon Lee., and Hyunkook Kahng, "MRED: A New Approach to Random Early Detection" 15th International Conference on Information Networking, February 2001

[14] S. Athuraliya, V. H. Li, S. H. Low, and Q. Yin, "REM: Active Queue Management," IEEE Network Magazine, Vol. 15, pp. 48–53, 2001.

[15. W. Feng, Apu Kapadia, and Sunil Thulasidasan,, "GREEN: Proactive Queue Management over a Best-Effort Network", IEEE GlobeCom (GLOBECOM 2002), Taipei, Taiwan, November 2002.

[16] W. Feng, D.D. Kandlur, D. Saha, and D. Saha, "The Blue active queue management algorithms," IEEE/ACM Transactions on Networking, 2002.

[17] J. Sun, K. Ko, G. Chen, S. Chan, and M. Zukerman, "PD-RED: To Improve the Performance of RED", IEEE Communication Letters, Vo. 7, No. 8, pp. 406-408, August 2003.

[18] C. Wag, Bin Liu, Y.Thomas Hou., and Kazem Sobraby., "LRED: A Robust Active Queue Management Scheme Based on Packet Loss Ratio", 23rd Annual Joint Conference on Performance, Computing and Communication 2004.

[19] L. Hu., and Ajay D.Kshemkalyani., "HRED: A simple and Efficient Active Queue Management Algorithm", 13th International Conference on Computer Communications and Networking ICCCN 2004, October 2004.

[20]Yue-Dong Xu., Zhen-Yu Wang., and Hua Wang., "ARED: A Novel Adaptive Congestion Controller", IEEE International Conference on Machine Learning and Cybernetics, August 2005.

[21] Shan Suthaharan, "Reduction of queue oscillation in the next generation Internet routers", Science Direct, Computer Communication, 2007.

[22] K. Chitra and G. Padamavathi, "Adaptive CHOKe: An Algorithm to Increase the fairness in Internet Routers," International Journal Advanced Networking and Applications, Vol. 01, Issue 06, pp. 282-386, 2010.

[23] K.Chitra and G.Padamavathi, "Classification and Performance of AQM-Based Schemes for Congestion Avoidance," International Journal of Computer Science and Information Security, Vol. 8, No. 1, 2010

[24] V. A. Reguera, F. F. Paliza, W. Godoy, and E. M. Fernandez, "On the impact of Active Queue management on VOIP quality of Service," Computer Communications, Vo. 31, pp. 73-87, 2008.

[25] I. Awan, S. Ahmed, and B. Ahmed, "Performance Analysis of Multimedia Based Web Traffic with QoS Constraints," Journal of Computer and System Sciences, Vol. 74, pp. 232-242, 2008.

[26] N. Selvam, and S. Radhakrishnan, "Processor Based Active Queue Management for Providing QoS in Multimedia Application," International Journal of Computer Science and Information Security, Vol. 7, No. 9, 2010.

[27] S. McCanne and S. Floyd, "ns Network Simulator", http://www.isi.edu/nsnam/ns.

[28] L. Breslau, et al., "Advanced in network simulation," IEEE Computer, Vol. 33, No. 5, pp. 59-67, May 2000.

[29] K. Fall and K. Varadhan, "The ns Manual," UC Berkeley, LBL, USC/ISI, and Xerox PARC, June 2003: http://www.isi.edu/nsnam/ns/doc/

**Gamal Attiya** graduated in 1993 and obtained his MSc degree in computer science and engineering from the Menufiya University, Egypt, in 1999. He received PhD degree in computer engineering from the University of Marne-La-Vallée, Paris-France, in 2004. He is currently Lecturer at the department of Computer Science and Engineering, Faculty of Electronic Engineering, Minoufiya University, Egypt. His main research interests include distributed computing, task allocation and scheduling, computer networks and protocols, congestion control, QoS, and multimedia networking.

**Heba A. El-Khobby** graduated in 1998 and received her M.Sc and PhD degree from the Tanta University, Egypt, in 2003 and 2009 respectively. She is currently a Lecturer at the Department of Electronics and Electrical Communications Engineering, Tanta University, Egypt. Her research interests are in the field of computer networks including congestion control, routing, mechanisms for resource management and QoS.

# An Improved Approach for Working outside the MANET by Extending MANET Routing Protocol

**Rashween Kaur Saluja**

**CS Department, RGPV, GGITS**
**Jabalpur, Madhya Pradesh, India**

**Abstract**

Mobile ad-hoc network have the attributes such as wireless connection, continuously changing topology, distributed operation and ease of deployment. We present a design space analysis of the problem of providing Internet connectivity for mobile ad hoc networks (MANETs). For widening the coverage area of the MANET there is a growing need to integrate these ad hoc networks to the Internet. For this purpose we need gateways which act as bridges between different protocols architectures. In this paper the AODV reactive routing protocol is extended to support the communication between the MANET and the Internet.

***Keywords:*** *Packet delivery fraction, Average end-to-end delay, Average throughput, routing overhead, Loss of data packets.*

## 1. Introduction

MANET is a collection of wireless mobile nodes that communicate with each other using multi-hop wireless links without any existing network infrastructure or centralized administration [12]. Each node in the network behaves as a router and forwards packets to other nodes. For several military and civil applications, networking the mobile or static nodes with wireless links in an ad hoc manner can be necessary and effective [1].

In this paper our goal is to design Internet connectivity for MANETs that can handle node mobility, both within and in between networks, having continuous and uninterrupted Internet connections whenever there is at least one potential route to one or more gateways.

To achieve this network interconnection, gateways that understand not only the IP suite, but also the MANET protocol stack, are needed. Thus, a gateway acts as a bridge between a MANET and the Internet and all communication between the two networks must pass through any of the gateways.

This paper evaluates approaches for gateway discovery. An interesting question is whether the configuration phase with the gateway should be initiated by the gateway (proactive method), the mobile node (reactive method) or by mixing these two approaches. All of them are based on the number of physical hops to gateway as the metric for the gateway selection.

When using proactive routing protocols, also called “table driven” protocols, mobile nodes continuously evaluate routes to all reachable nodes and attempt to maintain up-to-date routing information. The advantages of this type of protocols are discovery of the shortest path through network and availability of routes at the time of need, this reduces delays. The drawback of proactive routing protocols is providing a resistance to network topology changes.

On the other hand, when mobile nodes use reactive routing protocols, also called “on-demand” protocols, route discovery operation is performed only when a routing path is needed, and it is terminated when a route or no route has been found. A very important operation in reactive routing is route maintenance. The advantages of this type of protocols are efficiency, reliability and less control overhead. However, a major lack is a long delay caused by a route discovery operation in order to transmit data packets. Hybrid approach tries to combine the advantages of both. These protocols perform variously depending on type of traffic, number of nodes, rate of mobility, etc.

There are various mobility models such as Random Way Point, Reference Point Group Mobility Model (RPGM), Manhattan Mobility Model, Freeway Mobility Model, and Gauss Markov Mobility Model etc that have been proposed for evaluation [2], [5].

In this paper we have described the design and implementation of various gateway discovery approaches and studied the performance differentials of these approaches under different scenarios using ns2 based simulation.

The rest of the paper is organized as follows. Section 2 gives an overview of the related work so far. Section 3 describes the Routing Protocols for MANET whereas Section 4 analyzes the AODV routing protocol in detail. Section 5 describes the MANET Protocol Stack that supports AODV. Section 6 describes the integration of the

MANET and the Internet and the issues involved in MANET-Internet connectivity. The Enhanced AODV Routing Protocol which explains my work is described in Section 7. The different gateway discovery approaches are described in Section 8. The simulation setup and the network simulator-NS2 used are discussed in section 9. Also results are presented and analyzed in this section. Finally section 10 concludes the paper with future work.

## 2. Related Work

Mobile nodes in the Ad Hoc network need global addresses to communicate outside the MANET and node mobility should be properly dealt with [16][8]. Especially, when mobile nodes move to another area, their subnet changes and a new IP address must be obtained. Several solutions have been proposed to deal with the integration of MANETs to the Internet. Most of the proposed solutions require the addition of gateways and the routing protocols used within the Ad Hoc network. Since Internet gateways have two interfaces they are part of the Internet and the Ad Hoc network simultaneously. They understand the Internet protocol (IP) as well as a MANET routing protocol (e.g. AODV).

In this section we explore the most significant features of the main MANET interconnection mechanisms namely those from Wakikawa *et al [9]*, Jelger *et al*. [10], Singh *et al [11]* and Ros *et al.* Table I summarize the main features provided by each one.

Table 1
Summary of features of well known existing protocols.
P=Proactive, R= Reactive, H=Hybrid, A= Adaptive, RH=Routing Header, DR= Default Routing, OPT=Optional

| | Wakikawa | Jelger | Singh | Ros |
|---|---|---|---|---|
| GW Discovery | P/R | P | H | A |
| Multiple Prefix | Yes | Yes | No | Yes |
| Stateless/ful | less | less | n/a | less |
| DAD | Yes | No | n/a | Opt |
| Header/Default | RH | DR | Both | n/a |
| Limited Flooding | No | Yes | No | Yes |
| Load Balancing | No | No | Yes | No |
| Complete Spec. | Yes | Yes | No | Yes |

"Wakikawa" [9] defines two mechanisms, a reactive and a proactive one. In the reactive version, when a node requires global connectivity it issues a request message which is flooded throughout the MANET. When this request is received by a gateway, then it sends a message which creates reverse routes to the gateway on its way back to the originator.

Ratanchandani et al. [13] introduced a hybrid gateway discovery approach which combines the advantages of both the proactive and reactive approaches. This scheme uses AODV and two Mobile IP foreign agents for interconnecting the MANET with the Internet. The excessive flooding of the proactive approach is reduced by carefully controlling the TTL value of the foreign agent advertisement. This reduces the total number of hops that the advertisement can traverse. Thus only the mobile nodes close to the foreign agent receive the advertisement proactively. The nodes which are further away find the gateway following the reactive approach.

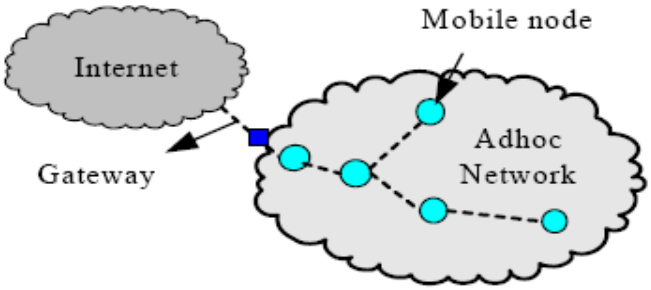



Fig. 1 Mobile Ad Hoc network connected with Internet

## 3. Routing Protocols for MANET

Routing protocols for MANETs can be broadly classified into three main categories:

### 3.1 Proactive routing protocols:

Every node in the network has one or more routes to any possible destination in its routing table at any given time.

### 3.2 Reactive routing protocols:

Every node in the network obtains a route to a destination on a demand fashion. Reactive protocols do not maintain up-to-date routes to any destination in the network and do not generally exchange any periodic control messages.

### 3.3 Hybrid routing protocols:

Every node acts reactively in the region close to its proximity and proactively outside of that region, or zone.

## 4. AODV (Adhoc on Demand Distance Vector)

There are two types of routing protocols which are reactive and proactive. In reactive routing protocols the routes are created only when source wants to send data to destination whereas proactive routing protocols are table driven. Being a reactive routing protocol AODV [14] uses traditional routing tables, one entry per destination and sequence numbers are used to determine whether routing information is up-to-date and to prevent routing loops. The maintenance of time-based states is an important feature of AODV which means that a routing entry which is not recently used is expired. The neighbors are notified in case of route breakage. Control messages used for the discovery and breakage of route are as follows:

### 4.1 Route Request (RREQ):

A route request packet is flooded through the network when a route is not available for the destination from source. The parameters are:

Table 2: Route Request Parameters

| Source Address | Request ID Source | Sequence Number | Destination Address | Destination Sequence Number | Hop Count |
|---|---|---|---|---|---|

A RREQ is identified by the pair source address and request ID, each time when the source node sends a new REQ and the request ID is incremented. After receiving of request message, each node checks the request ID and source address pair. The new RREQ is discarded if there is already RREQ packet with same pair of parameters. A node that has no route entry for the destination, it rebroadcasts the RREQ with incremented hop count parameter.

### 4.2 Route Reply (RREP):

Once find out the valid route to the destination or if the node is destination, a RREP message is sent to the source by the node.
The following parameters are contained in the route reply message:

Table 3: Route Reply Parameters

| Source Address | Destination Address | Destination Sequence Number | Hop Count | Life Time |
|---|---|---|---|---|

### 4.3 Route Error Message (RERR):

The neighborhood nodes are monitored. When a route that is active is lost, the neighborhood nodes are notified by route error message (RERR) on both sides of link.

### 4.4 Hello Messages:

The HELLO messages are broadcasted in order to know neighborhood nodes. The neighborhood nodes are directly communicated. In AODV, HELLO messages are broadcasted in order to inform the neighbors about the activation of the link. These messages are not broadcasted because of short time to live (TTL) with a value equal to one.
Route discovery process begins when one of the nodes wants to send packets. That node sends Route Request (RREQ) packets to its neighbors. Neighbors return RREP packets if they have a corresponding route to destination. However, if they don't have a corresponding route, they forward RREQ packets to their neighbors, except the origin node. Also, they use these packets to build reverse paths to the source node. This process occurs until a route has been found. The algorithm uses hello messages. If hello messages stop coming from a particular node, the neighbor can assume that the node has moved away and mark that link to the node as broken and notify the affected set of nodes by sending a link failure notification (a special RREP) to that set of nodes. These messages are broadcasted because with TTL value equal to one.
When a source node does not have routing information about destination, the process of the discovery of the route starts for a node with which source wants to communicate. The process is initiated by broadcasting of RREQ. On receiving RREP message, the route is established. If multiple RREP messages with different routes are received then routing information is updated with RREP message of greater sequence number. If the originator node does not receive a RREP message within a certain time interval, it exponentially increments the time interval and increases the diameter of the searching ring.
In conclusion, the simple design, the low routing overhead and the ring searching technique make AODV an attractive solution for networks in which the available bandwidth is limited and nodes can form organized groups.

## 5. MANET Protocol Stack

Figure 2 shows the protocol stack of MANET which consists of five layers: physical layer, data link layer, network layer, transport layer and application layer. It has similarities to the TCP/IP protocol suite. As can be seen, the OSI model's session, presentation and application layers are merged into one section, the application layer in MANET and TCP/IP suite.

OSI is a layered framework for the design of network systems that allows for communication across all types of computer systems. Because TCP/IP was designed before the OSI model, its layers do not correspond exactly to the OSI layers. The lower four layers are the same in both models but the fifth layer in the TCP/IP suite (the application layer) is equivalent to the combined session, presentation and application layers of the OSI model.
The main difference between MANET and TCP/IP suite protocol stacks lies in the network layer. MNs (which are both hosts and routers) use an ad hoc routing protocol to route packets. In the physical and data link layer, MNs run protocols that have been designed for wireless channels. In this paper work, the standard IEEE 802.11 is used as simulation tool.

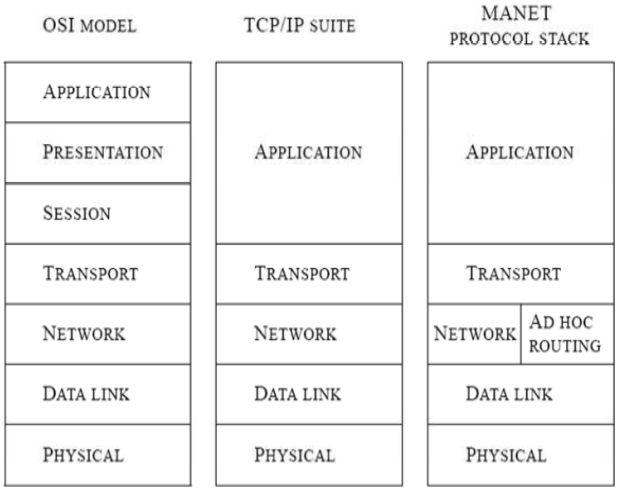



Figure 2: OSI Model, TCP/IP Protocol Suite and MANET Protocol Stack

When extended AODV MANET routing protocol is considered, the network layer is divided into two parts: The fixed network and Ad Hoc Routing in the MANET. The protocol used in the fixed network part is Internet Protocol (IP) and the protocol used in the ad hoc routing part is AODV.
In the transport layer, the User Datagram Protocol (UDP) is used in this work. The Transmission Control Protocol (TCP) is not used because different research works revealed that, TCP does not perform well in MANETs. This is because of the fact that, in wired networks, lost packets are almost always due to congestion but in MANETs, lost packets are more often caused by other reasons like link breakage due to mobility or interference [15].

## 6. Connectivity of MANET with Internet

Whenever a MN is to send packets to a fixed network, it must transmit the packets to a GW [3]. The protocol stacks involved during communication between a MANET and the fixed Internet node is shown in Figure 3. A GW acts as a bridge (not the network device) between a MANET and the Internet. Therefore, it has to implement both the MANET protocol stack and the TCP/IP suite.

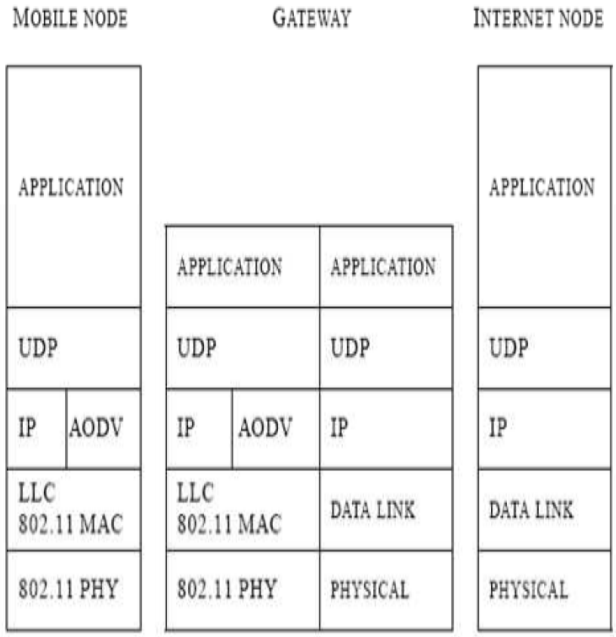



Figure 3: The Protocol Stack Used By Mobile Node, Internet Node and Gateway

## 7. Enhanced AODV Protocol

The enhanced AODV MANET routing protocol to support the three types of GWDAs.

### 7.1 The Enhanced Route Request

The enhanced RREQ message contains exactly the same fields with the same functions as the ordinary RREQ message, except for a flag as shown in Figure 4. This flag is called 'Internet-Global Address Resolution Flag' and is referred to as the I-flag. The, I-flag is used for global address resolution. It indicates that the source node requests global connectivity. The RREQ_I message plays the same role as the router solicitation message of ICMP. The RREQ_I message is used to reactively discover a gateway.

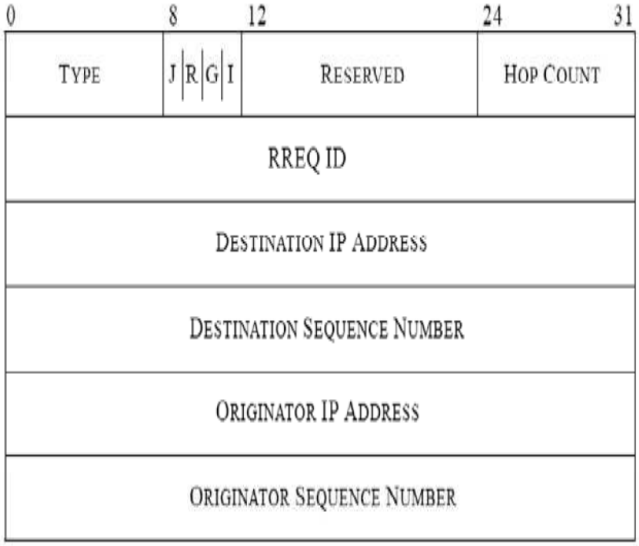



Figure 4: Enhanced Route Request Message Format

## 7.2. The Enhanced Route Reply

The enhanced RREP message contains exactly the same fields with the same functions as the ordinary RREP message, except for a flag. The RREP message is similarly extended by the Internet Global Address Resolution Flag or the I-flag. The RREP message extended with the I-flag is known as RREP_I message. This flag is used for global address resolution. It indicates that the gateway information is carried by the RREP_I message. The RREP_I message plays the same role as the router advertisement message of ICMP.

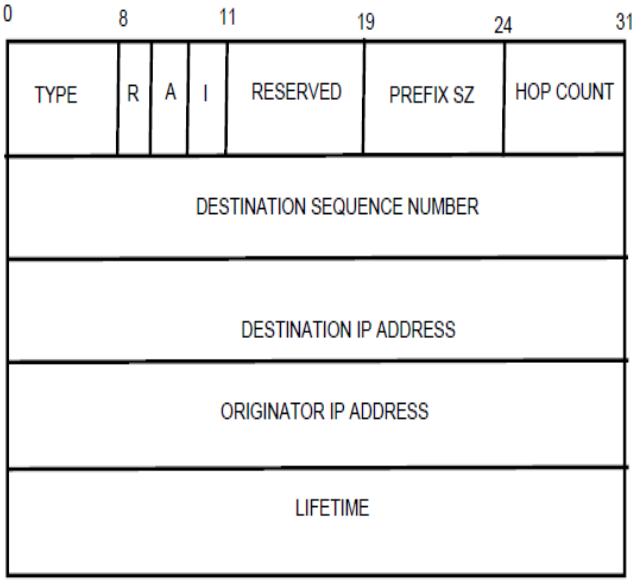



Figure 5: The Enhanced Route Reply Message Format

## 7.3. The Gateway Advertisement (GWADV)

GWADV is approximately a RREP_I message but it is extended to have a GWADV_ID, just like the RREQ ID of the RREQ packet in AODV MANET routing protocol. The GWADV_ID helps to avoid duplicated advertisement messages. When a MN receives a GWADV, it first checks to determine whether a GWADV with the same originator IP address and GWADV_ID already have been received during the last broadcast ID save seconds. If such a GWADV message has not been received, the message is rebroadcasted. Otherwise, if received, the newly received GWADV is discarded. Hence, duplicated GWADVs are not forwarded.

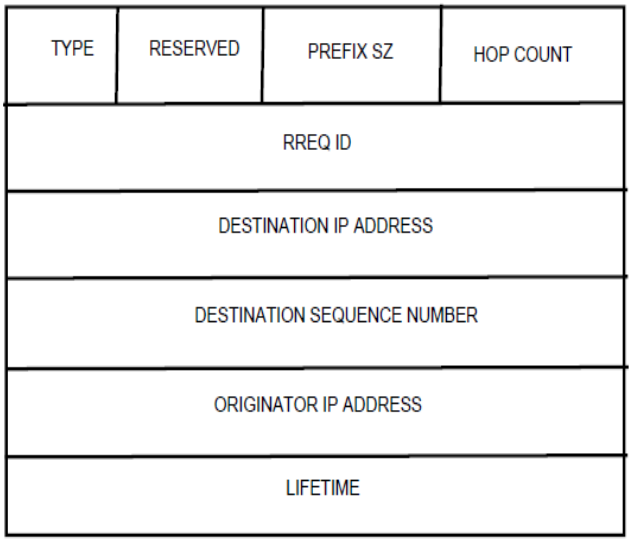



Figure 6: GWADV Message Format

## 7.4. The Default Route (Route to the Gateway)

The MN needs a route to a gateway, which it uses as its default route to send packets to the Internet. This GW information can be obtained in three different ways. One option is to rely on periodically advertised messages from the gateway (GWADVs), or by sending a RREQ_I to the ALL_MANET_GW_MULTICAST address (i.e. by sending to the GW nodes' group address). There is also a third option, for the sake of updating the default route entry, the GW nodes are made to reply RREQ messages with RREP_I messages, as a result, a MN can get default route by sending RREQ message to the gateway. However, this happens only when a MN is performing radial ring search before it gets the information, whether the destination node is within the ad hoc network or in the fixed network.

## 7.5 Gateway Operation upon Reception of RREQs

When a gateway receives a RREQ, it looks in its routing table searching for the destination IP address specified in the RREQ message. If the address is not found in the routing table, the gateway has to send a RREP_I back to the originator of the RREQ. On the other hand, if the gateway finds the host route in its routing table, it should not unicast back a RREP_I to the originator of the RREQ "because the destination is then assumed to be inside the MANET". Also "A gateway replies every received RREQ with a RREP_I."

## 8. Gateway Discovery

There are three types of GWDAs depending on the GW configuration phase initiation and also on the method of route update. If the configuration phase is initiated by the gateway, proactive method is used. But, if the initiation is made by a MN, reactive method is used. The combination of these two methods is called hybrid proactive/reactive method. The basic difference between the algorithms is highlighted below.

### 8.1 Proactive Gateway Discovery

The proactive GW discovery is initiated by the GW itself. The GW periodically broadcasts a GWADV message which is transmitted after expiration of the gateway's advertisement interval timer that is the time between two consecutive advertisements must be chosen with care so that the network is not flooded unnecessarily. All MNs residing in the gateway's transmission range receive the advertisement.
Upon receipt of the advertisement, the MNs that do not have a route to the GW create a route entry for it in their routing tables. MNs that already have a route to the GW update their route entry for the gateway. Next, the advertisement is forwarded by the MNs to other MNs residing in their transmission range. To assure that all MNs within the MANET receive the GW advertisement. The number of retransmissions is determined by network diameter.
However, this will lead to enormously many unnecessary duplicated advertisements. This is disadvantage. Limited resources in a MANET, such as power and bandwidth, will be excessively used.

### 8.2 Reactive Gateway Discovery

The reactive GW discovery is initiated by a MN that is to initialize or update information about the gateway. The MN broadcasts a RREQ_I to IP address for the group of all gateways in a MANET. Thus, only gateways are addressed by this message and only they process it. Intermediate MNs that receive the message just forward it by broadcasting it again. Since the message format is RREQ, which has a unique request id field duplicated RREQ_Is are discarded. Upon receipt of a RREQ_I, a GW unicasts back a RREP_I which, among other things, contains the IP address of the gateway.
The advantage of this approach is that RREQ_Is are sent only when a MN needs the information about reachable gateways. Hence, periodic flooding of the complete MANET, which has obvious disadvantage, is prevented.

### 8.3. Hybrid Gateway Discovery

To minimize the disadvantages of proactive and reactive gateway discovery, the two approaches can be combined. This results in a hybrid proactive/reactive method for gateway discovery. For mobile nodes in a certain range around a gateway, proactive gateway discovery is used. Mobile nodes residing outside this range use reactive gateway discovery to obtain information about the gateway.
The gateway periodically broadcasts a RREP_I message (see Figure 5) which is transmitted after expiration of the gateway's timer, ADVERTISEMENT_INTERVAL (see Table 5). All mobile nodes residing in the gateway's transmission range receive the RREP_I. Upon receipt of the message, the mobile nodes that do not have a route to the gateway create a route entry for it in their routing tables. Mobile nodes that already have a route to the gateway update their route entry for the gateway. Next, the RREP_I is forwarded by the mobile nodes to other mobile nodes residing in their transmission range. The maximal number of hops a RREP_I can move through the mobile ad hoc network is ADVERTISEMENT_ZONE (see Table 5). This value defines the range within which proactive gateway discovery is used.
When a mobile node residing outside this range needs gateway information, it broadcasts a RREQ_I to the ALL_MANET_GW_MULTICAST address. Mobile nodes receiving the RREQ_I just rebroadcast it. Upon receipt of this RREQ_I, the gateway unicasts back a RREP_I.

## 9. Network Simulator (NS2)

Network Simulator 2(NS2), is a discrete event NS. The University of California at Berkeley and the VINT project [4] has developed it. It is popular for its extensibility (due to its open source model). NS2 is popularly used in the simulation of routing and multicast protocols, among others, and is heavily used in researches based on ad hoc networks. NS supports an array of popular network protocols, offering simulation results for wired and wireless networks.

NS2 supports system programming language C++ for detail implementation and scripting language TCL for configuring and experimenting with different parameters quickly. NS-2 has all the essential features. NS is written in C++, with an OTcl interpreter. The C++ part, which is fast to run but slower to change, is used for detailed protocol implementation. The OTcl part, on the other hand, which runs much slower but can be changed very quickly, is used for simulation configuration.

### 9.1 Simulation Scenario

In order to evaluate the performance of the three gateway discovery methods, I used the network simulator ns-2 (ns-2.31).
First, the source code of AODV in ns-2 was extended to provide access to mobile stations. Then the three gateway discovery methods were implemented.
The simulations were conducted on an Intel(R) Core™ i3 CPU processor at 2.40 GHz, 3 GB of RAM running cygwin in Windows XP.
The mobile nodes move according to the "random waypoint" model. The movement patterns are generated by CBR's movement generator (setdest).
The traffic connection pattern is generated by CBR traffic generator (cbrgen.tcl).

### 9.2 Simulation Environment

The Simulation environment is setup, by placing two GW nodes, which are fixed and are connected to two routers on the fixed network. Each router is connected to a host in the fixed network. The routers are also connected to each other to facilitate routing from any GW to any host in the fixed network. The GW nodes are located at (150,300) and (850,300).
At the third layer i.e. the network layer, extended AODV is used as the ad hoc routing protocol, whereas DCF is used at the MAC sub layer with its default values for the contention parameters. Finally, at the physical stations use IEEE 802.11 DSSS.
The parameters that are common for all simulations are given in Table 4 and the parameters that are specific for some simulations are shown in Table 5.

Table 4: Parameters for Simulation

| Parameter | Value |
|---|---|
| Transmission Range | 250 m |
| Simulation Time | 1000 s |
| Simulation Area | 1000 × 700 |
| Number of Mobile Nodes | 6,12,18,25 mobile nodes |
| Number of Sources | 1 |
| Number of Gateways | 2 |
| Traffic Type | CBR |
| Packet Rate | 5 packets/s |
| Packet Size | 512 bytes |
| Pause Time | 2 s |
| Maximum Speed | 10 m/s |

Table 5: Specific Parameters Used in Some Simulations

| Parameter | Value |
|---|---|
| ADVERTISEMENT_INTERVAL | 5 Seconds |
| ADVERTISEMENT_ZONE | 3 Hops |

ADVERTISEMENT_INTERVAL is used when proactive and hybrid discovery methods are used.
ADVERTISEMENT_ZONE is used for hybrid gateway discovery method and defines the range within which proactive gateway discovery is used.

### 9.3 Screenshot

A screenshot of the simulation scenario is shown in Figure 7. The eighteen mobile nodes that are marked with a ring are the sources. The two hexagonal nodes are the gateways and the four square nodes are the two hosts and the two routers.

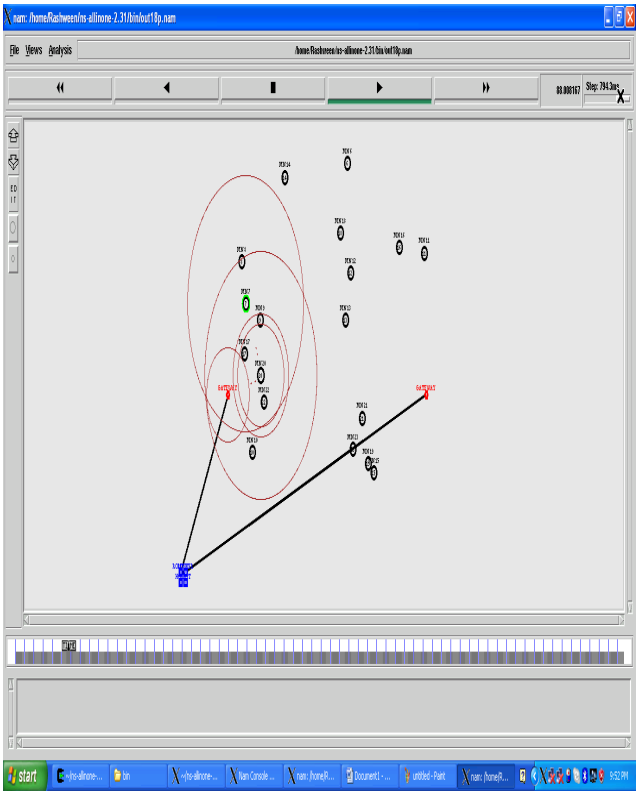


Figure 7: Screenshot of Simulation Environment

## 9.4 Simulation Results

### 9.4.1. Packet Delivery Ratio V/s Number of Nodes:

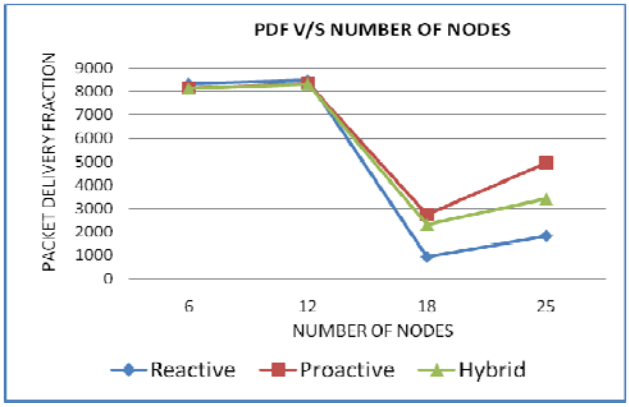



Figure 8: Packet Delivery Fraction V/S Number of Nodes

From the simulation results we see that the proactive approach has better packet delivery performance than the reactive approach. This happens because - due to the periodic update of route information from the gateway, routes from all the nodes to the gateway are always available. As a result majority of the packets are delivered smoothly. In case of reactive approach, a node wishing to send data to the destination needs to find the route to the gateway first. This takes a certain amount of time and no packet can be sent during this period due to the unavailability of routes.

Moreover, in case of proactive approach, due to regular exchange of gateway information, routes are always optimized and the nodes have fresher and shorter routes to the destination. This reduces the chances of link breaks and increases the packet delivery ratio.

### 9.4.2 Average End To End Delay V/s Number of Nodes:

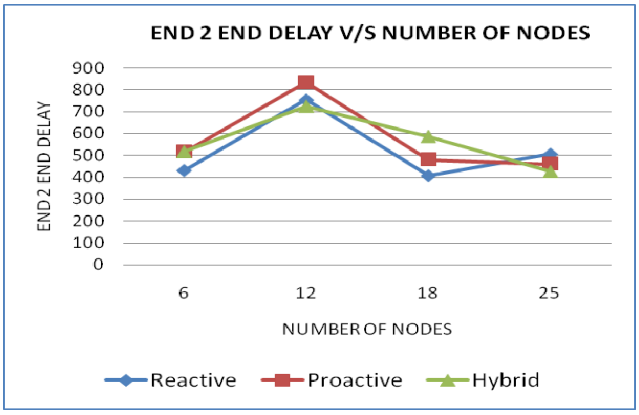



Figure 9: End To End Delay V/S Number of Nodes

In terms of the average end-to-end delay, the delay for reactive and hybrid gateway discovery approaches is much less as compared to the proactive gateway discovery when we increase the number of nodes. When the number of nodes is less, all the approaches suffer from greater average end-to-end delay.

### 9.4.3. Average Throughput V/s Number of Nodes

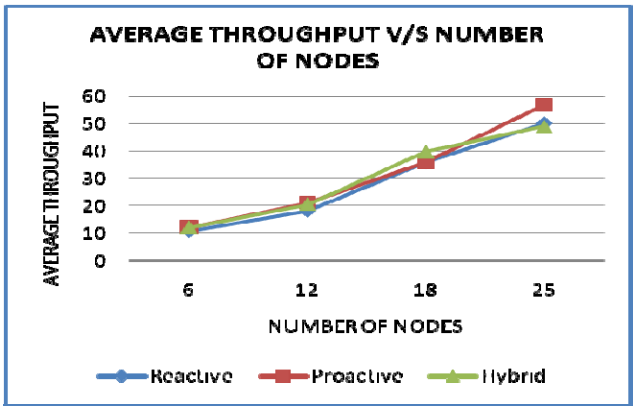



Figure 10: Average Throughput V/s Number of Nodes

As far as average throughput is concerned, when the number of nodes is less, then all three approaches have

almost same throughput. But when the number of nodes is increased, the proactive approach outperforms than the reactive and hybrid approaches.

9.4.4. Routing Overhead V/s Number of Nodes:

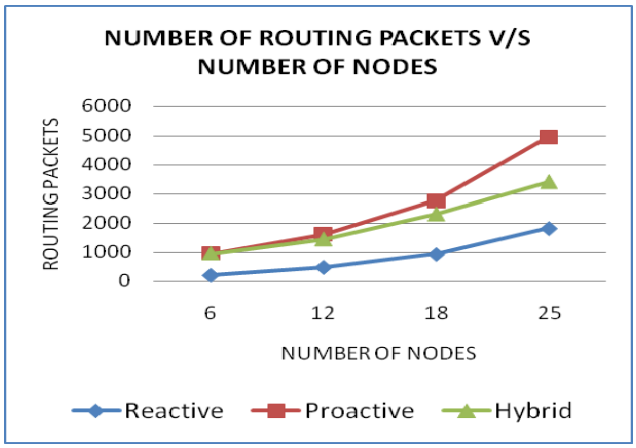



Figure 11: Number of routing packets v/s Number of Nodes

In case of routing packets, the proactive approach clearly outperforms reactive and hybrid approaches.

9.4.5. Loss of Packets V/s Number of Packets:

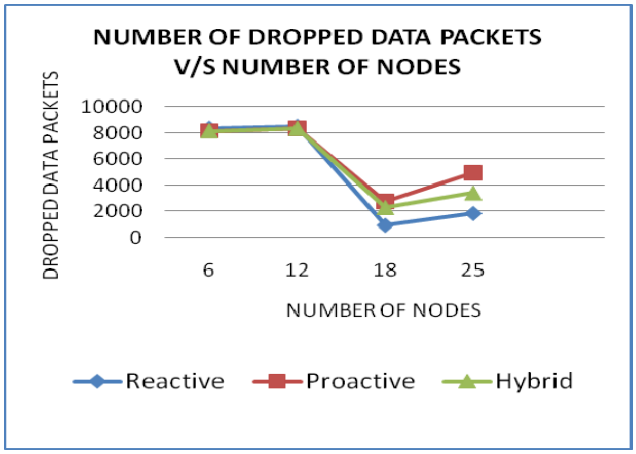



Figure 12: Number of Dropped Data Packets V/s Number of Nodes

For number of dropped data packets, when number of nodes is less, all three approaches remains almost constant. With more number of nodes, the number of dropped data packets increases for the proactive approach because congestion increases. Whereas for the reactive approach, with increasing number of nodes, the number of dropped data packets, decreases because it sends packets only when there is a need. The hybrid approach being a combination of proactive and reactive approaches, its number of dropped data packets lies between them.

## 10. Conclusions

In the paper, MANET routing protocol-AODV has been extended to route packets, not only within a MANET but also between a wireless MANET and the wired network. The communication between the wireless and the wired network must pass through these nodes, which are referred to as gateways. In this thesis work, three methods for detection of these gateways have been presented, implemented and compared.

The three methods for gateway detection are referred to as reactive, proactive and hybrid gateway discovery. The comparison between these methods provides us useful information.

Regarding the packet delivery ratio in proactive gateway discovery approach, due to regular exchange of gateway information, routes are always optimized and the nodes have fresher and shorter routes to the destination. This reduces the chances of link breaks and increases the packet delivery ratio. On the other hand in reactive approach, a node continues to use a longer route until it is broken even if an alternate shorter route is available. This reduces the packet delivery fraction. The packet delivery performance of the hybrid approach falls between that of the proactive and reactive approaches.

In terms of the average end-to-end delay, the reactive gateway discovery suffers from less average end-to-end delays compared to proactive and hybrid gateway discovery approach.

As far as average throughput is concerned, initially, when the number of nodes is less, all three approaches have almost same throughput. But when the number of nodes is increased, the proactive approach performs better than the reactive and hybrid approaches.

In case of routing packets, the proactive approach clearly outperforms reactive and hybrid approaches.

For number of dropped data packets, when number of nodes is less, the reactive approach has more number of dropped data packets. With more number of nodes, the number of dropped data packets increases for the proactive approach because congestion increases.

## References

[1] M. Frodigh, P. Johansson, P. Larsson. Wireless ad hoc networking—the art of networking without a network. Ericsson Review No. 4, 2000.

[2] Tracy Camp, Jeff Boleng, Vanessa Davies “A Survey of Mobility Models for Ad Hoc Network Research”, Wireless Communication & Mobile Computing (WCMC): vol. 2, no. 5, pp. 483-502, 2002.

[3] M.Rosenschon et al. "Gateway Discovery Algorithm for Ad-Hoc Networks using HELLO Messages". IWWAN 2005, London, May 2005.

[4] Charles E. Perkins and Pravin Bhagwat. The “ns Manual”, A Collaboration between researchers at UC Berkeley, LBL, USC/ISI, and Xerox PARC, December 19, 2008
[5] N.Aschenbruck, E.Gerhands-Padilla, P.Martini,” A Survey on mobility models for Performance analysis in Tactical Mobile networks,” Journal of Telecommunication and Information Technology, Vol.2 pp.54-61, 2008
[6] Rakesh Kumar, Anil K. Sarje and Manoj Misra“Review Strategies and Analysis of Mobile Ad Hoc Network- Internet Integration Solutions”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 6, July 2010.
[7] Koushik Majumder et al. “Implementation and Performance Evaluation of the Gateway Discovery Approaches in the Integrated MANET Internet Scenario”. International Journal on Computer Science and Engineering (IJCSE), Kolkata, India, Vol. 3 No. 3 Mar 2011.
[8] Y. Sun, E.M. Belding_Royer, C.E. Perkins,” Internet Connectivity for Ad Hoc Mobile Networks,” International Journal of Wireless information Networks, Special Issue on Mobile Ad Hoc networks (MANETs): Standards, Research, Applications 9 (2) (2002) 75-88.
[9] R. Wakikawa, J. Malinen, C. Perkins, A. Nilsson, and A. Tuominen, “Global Connectivity for IPv6 Mobile Ad Hoc Networks,” Internet-Draft “draft-wakikawa-manet-globalv6-03.txt”. Oct. 2003.
[10] C. Jelger, T. Noel, and A. Frey, “Gateway an Address Auto configuration for IPv6 Ad Hoc Networks,” Internet- Draft “draft-jelger-manet-gateway-autoconf-v6-02.txt”. Apr. 2004.
[11] S. Singh, J. Kim, Y. Choi, K. Kang, and Y. Roh, “Mobile Multi-gateway Support for IPv6 Mobile Ad Hoc Networks,” Internet-Draft, “draft-sinhg-manet-mmg-00.txt”, June 2004.
[12] S. Shah, et al., “Performance Evaluation of Ad Hoc Routing Protocols Using NS2 Simulation,” Proceedings of the National Conference on Mobile and Pervasive Computing (CoMPC-2008), Chennai, India, August 2008.
[13] P. Ratanchandani and R. Kravets. “A Hybrid Approach to Internet Connectivity for Mobile Ad Hoc Networks”, Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), New Orleans, Louisiana, USA, 16-20 March, 2003
[14] Kumar, R., Misra, M. and Sarje, A.K. (2007) “An Efficient Gateway Discovery in Ad Hoc Networks for Internet Connectivity”, Proc. of the International Conference on Computational Intelligence and Multimedia Applications, pp 275-281.
[15] Holland G.; Vaidya N. Analysis of TCP Performance over Mobile Ad Hoc Networks, in Proceedings of IEEE/ACM MOBICOM, 1999.
[16] U. Jonsson, F. Alriksson, T. Larsson, P. Johansson, and G.M. Maquire, “MIPMANET: Mobile IP for Mobile Ad Hoc Networks,” Proceedings of IEEE/ACM Workshop on Mobile and Ad Hoc Networking and Computing (MobiHoc 2000), Boston, MA USA, pp. 75-80, August 1999.

**Rashween Kaur Saluja** has received her BE degree- from HCET in 2008 and is pursuing M-Tech from SRIT. Published papers in CCCT conference (Greater Noida Chapter) on 7-8 August, 2011 and in “International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 (Print), VOL-I journal. Also published paper in IEEE sponsored Second International Conference on Computer Communication and Informatics (ICCCI 2012) to be held during 10-12, January 2012, Coimbatore. Has keen interest in mobile adhoc networks and MANET Internet-Integration and Performance Evaluation.

# Pattern Design for Software Testing Base Finit Automato Machines

**Seyyede Roya Alavi**[1]

[1]**Department of Computer, Khoy Branch, Islamic Azad University, Khoy, Iran**

**Abstract**

On method for effective testing is the identification of critical rout in the program. Standardized test of software is somehow impossible because the production and control of critical routs is difficult for software with average size. For creating control in the routs Finit Automate Machines (FAM) are used, in order to design a series of grammer and then the language for each rout by making use of FAM. Grammers create a chain sequences which should be followed through the program. These produced sequences decrease the complexity of identification of errors for effective testing in the process of examination purposefully.

***Keywords:*** *Finit Automato Mach, Grammer, Path, Software Testing.*

## 1. Introduction

The process of testing any software system is an enormous task which is time consuming and costly software testing spends almost 50% of software system development resources. Random test simply runs the inputs and then clarifies the performed structures but it can't extract some of the accessible information from black box. Dynamic test white and black box methods are used in combination to produce finite amount of test instances. Testing involves three main steps: generation a set of test inputs, execution thos inputs on the program under tests, and then checking whether the test executions reveal faults.

Software testing is a time consuming and costing process. For applying a standardized test all possible paths should be studied. As the program follows running paths according to different inputs, it is impossible to explore all the errors of the program before its practicle use by the software users. Dealing all run paths before program delivery is very difficult or maybe impossible. So, many of errors are concealed in the program and will be recovered after using.An Automatic testing software can generally decrease the cost of software development. Moreover, it can causes quick running of the test and the reliability of test result decreases. Automatic testing is not a direct and confort a progress process [1].

Error Location finding methods are generally distinguished into two types: static and dynamic. Static methods tries to identify the location of errors in the program according to Program Dependent Graph [2,3,4]. On the other hand, dynamic methods try to approximate location of the error by comparing successful and unsuccessful runs of the program [5,6]. Gathering needed data for modeling of run paths of program is an important problem in error finding of the software. Storing all run data, which is produced by the program is not possible in the practice. But only a part of run data of program can be stored.

Solving this problem a pattern of administration can be offered. According to this pattern the behavior of the program can be analyzed during different runs. Offering the pattern for modeling of run paths of the program can be done by grammers in FAM.

The rest of the paper is arranged as follow: In section 2, the review of literature of software testing are discussed. In section 3, suggested methods, by the use of Finit Automate Machines are examined and in section 4, conclusion is given.

## 2. Related Work

Static testing methods running the program on test by input data testing and recovering its output [7]. The goal of dynamic analyzing methods is comparing the behavior of running program time in successful and unsuccessful run, which detection the program errors [5]. Dynamic methods applying with care of successful and unsuccessful data and without any attention to the programs static structure. Static method is distinguished into two type: black box and white box. Black box testing is essential only in examining the output in response to the input data.

White box testing is made by the use of information about how it works inside the unit [8,9,10]. In dynamic techniques the data results of run program is stored and after analyzing them chaos inside the program code is explored and introduced to the user. Previous research for finding the location of errors from timing behavior of program administration such as program spectra used memory graphs and history of examining determinants of the program [11,12]. Among analyzing technique of dynamic method, techniques based on examining of determinant for finding locations of errors were more successful [13,14]. In dependent graph program, data dependence and controlling dependence of program are shown [15]. In fact, the nodes of this graph are sentences of the program and determinant statements. And their edge shows data dependence and controlling dependence of the program.

Similar to control and data flow coverage criteria, state base testing relies on coverage criteria defined . These include: all-transitions, all-transition-pairs, all predicates and allsequences. Other coverage criteria like all-states and all-ntransitions are very often used. However, all-states (each state must reached at least once) is subsumed by all-transitions criterion which is in turn subsumed by all-n-transitions. While these criteria are typical for state-based testing, in many research publications state machines have been extended to deal with dataflow coverage criteria [15]. In Fig.1, the code of one program with related CFG (Control Flow Graph) is illustrated.

```
0.  Function (int a, int b)
1.  {
2.  Int c;
3.  If (b>a) {
4.  c=a;
5.  a=b;
6.  b=c;
7.  }
8.  C=a/b;
9.  While (c!=0)
10. {
11. a=b;
12. b=c;
13. c=a/b;
14. }
15. Return  b;
16. }
```

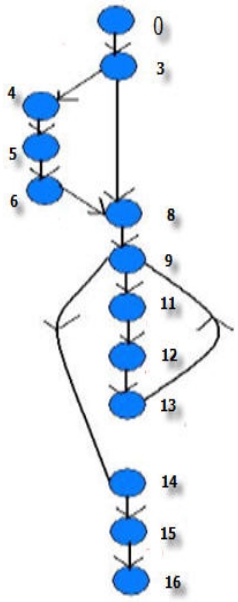



Fig.1 Code With CFG

## 3. Proposed Pattern

The path which the programs starts and ends, is the program behavior. It is clear that a program can have different behaviors. The more complexity of software the more behavior domains of the program. The behavior of the program can be sequence of its occurrences, such that, the occurrence can be the calling function, running a line of program, or the return value of function. The behavior of program can be true or false. LTL (Linear Temporal Logic) formulae can be converted mechanically into testautomata and then used in a model checking procedure, using the algorithm outlined [16]. The automaton will accept all those, and only those, execution sequences that correspond to a violation of the property [16]. The model checker SPIN contains the conversion algorithm, and can detect the violating sequences with a standard model checking run [16].

Any violations that are detected can then be reported as execution traces through the original implementation source code of the application.

The test automata are often also simple enough that they can be constructed by hand, and in some cases the hand-tuned automata are smaller than the machine generated ones,which translates to reduced run-time requirements for the model checking process [16].

### 3.1 Finit Automato Machine

Machines are design according to a language, and an input string enter finally gives an output which are YES or NO. In this part, the program characteristics are enters instead of language characteristics and the output is produced based on the language of machine grammer. The figure of a machine is illustrated in Fig. 2. The finite Automate which display with M, is formed from five elements: M = ( Q, ∑, S, $q_0$, F), in which Q is series of $q_0$,$q_1$,$q_2$, …, $q_i$ , ∑ is series of input string scripts, S is the rules of transfer or shifts, q0 is a member of Q and F is series of final conditions.

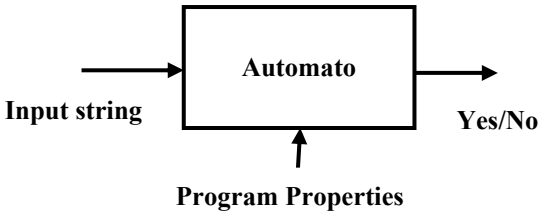



Fig.2 Automato

There is used a graph for showing a finit automate machine, Fig.3 display a simple graph of FAM. Such that the communication between state is called edge and illustrated S(qi , a) = qj.

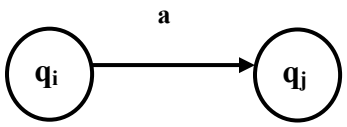



Fig. 3 edge of Automato

The grammer is used for description the language of FAM. The grammer is formed of four elements which expressing with G: G = (V, T, S, P). In grammer G, V element implies the de-terminal series and T implies the terminals, S is the sign of start and P is sign for series of theorems. For obtain the sequence of strings carring the left side of string and right side of string are replaced in a term, which said derivation.

***Grammer:***

***S→ aSb***
***S→ λ***

For example, the sequence of ab, aabb is done by following:

*Derivation:*
*S→ aSb → ab*
*S→ aSb→ aaSbb → aabb*

The path which is negotiate during running, is designed with FAM. Using FAM cause designing a series of grammer for each function inside the testing program. In Fig. 4 a FMA gas designed for program in Fig. 1. For drawing a graph of this FAM, follows that each line of program numbered and each line is a state and edges of graph for defining function take F and "RETURN", data defining, "{", "}" lines take label and because of conditions, like "IF", are labeled I and because of "WHILE", are labeled W, because of "ELSE" labeled E, because of correcting condition labeled C inside the while, if, … and because of correct and wrong conditions it gives t and F. because of every rule and terms which are running in the program, the label S is given for edge of graph.

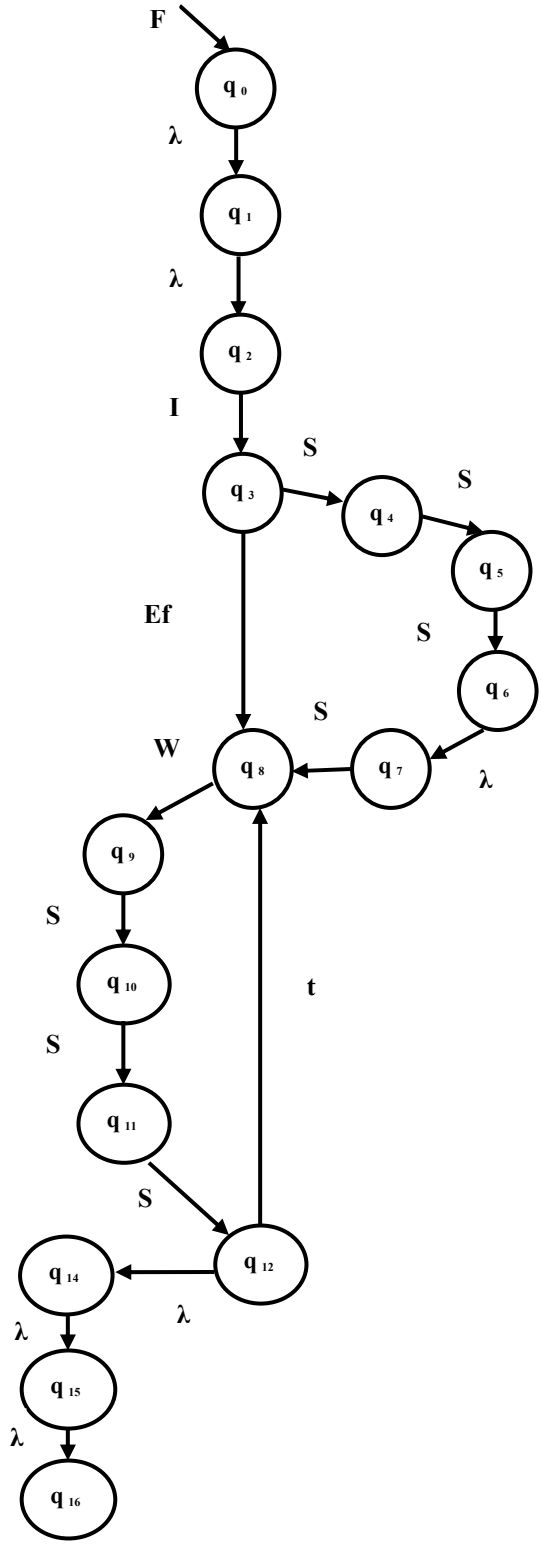



Fig. 4 Finit Automato Machine

In this case, the value of de-terminal V element and value of terminal T element are given:

V = { F, E, I, W, C, S}
T={ S, λ, t, f}

Therefore, the grammer which applied for Fig. 4 4 is in the followed. According to this grammer, language (like L) developed for it.

*Grammer Fig. 4:*

$F \rightarrow IF \mid EF \mid \lambda$
$I \rightarrow CS \mid \lambda$
$C \rightarrow t \mid Ef$
$S \rightarrow sS \mid s \mid \lambda$
$E \rightarrow SW \mid \lambda$
$W \rightarrow CW \mid S \mid \lambda$

*Language:*

$L=\{ (fs)^n (tsss)^m \mid n >= 0, m > 0\}$

Different strings are designed for grammers which displays the sequence of followed paths, next these sequences compared with that language. If the sequence of produced string is differ from produced strings of the languages, there is an error in the code. For example, if there is a wrong in line 3 (b<a) string "tsss" for input (2,3) produced from grammer and language. String "fsf" produced from grammer and shows that inside of part "IF", because of wrong conditions, deviated from run path of program.

*Derivation(true by grammer and language):*

$F \rightarrow IF \rightarrow I \rightarrow CS \rightarrow tS \rightarrow tsS \rightarrow tssS \rightarrow tsss$

*Derivatin (false by grammer):*

$F \rightarrow IF \rightarrow I \rightarrow CS \rightarrow fES \rightarrow fSWS \rightarrow fsCWS \rightarrow fsfEWS \rightarrow fsfWS \rightarrow fsfS \rightarrow fsf$

## 4. Conclusion

In software testing, location finding or suspect cases for errors in codes of program is the goal. FAM is used for finding the errors location, so that one way for error detection designed base on grammer and language is for program. In this paper, using this pattern, all the paths should followed are designed used of program language. All these produced series should be the string series which produced of program grammer. In this case, the left series of strings are recording as paths which the errors occur in them.

## References

[1] Srivastava, P. and Kim, T., "Application of Genetic Algorithm in Software Testing", International Journal of software engineering and its applications vol.3.no.4,2009, pp. 87-95.
[2] W. R. Bush, J. D. Pincus and D. J. Sielaff, "A Static Analyzer for Finding Dynamic Programming Errors", Software Practice and Experience, Vol. 30, No. 7, 2000, pp. 775−802.
[3] D. L. Detlefs, R. M. Leino, G. Nelson, J. B. Saxe, "Extended Static Checking", SRC Research Reports SRC−159, Compaq SRC, December 1998.
[4] W. E. Wong, S.S. Gokhale, and J.R. Horgan, "Measuring distance between program features", In International Conference of Computer Sotware and Applications (COMPSAC), 2002, pp. 307-312.
[5] T. Ball,"The Concept of Dynamic Analysis", In Proceedings of the 7th European Software Engineering Conference and the7th ACM SIGSOFT Symposium on Foundations of Software Engineering (ESEC/FSE'99), September 1999, pp.216-234.
[6] M.D. Ernst, J. Cockrell, W.G. Griswold and D. Notkin, "Dynamically Discovering Likely Program Invariants to Support Program Evolution", In IEEE Transactions on Software Engineering,Vol. 27, No. 2, February 2001.
[7] Myers, G. J., The Art of Software Testing, Revised and Updated by Tom Badgett and Todd M.Thomas with Corey Sandler,John Wiley & Sons , Inc, Second Edition, 2004.
[8] BCS SIGIST, "Standard for Software Component Testing", British Computer society, SIGIST, 2001.
[9] Gardner, D., "Software Testing Guide", Information management systems& services, California institute of technology, 2006.
[10] B. Liblit, M. Naik, A.X. Zheng, A. Aiken, and M.I. Jordan, "scalable statistical bug language Design and Implemenation (PLDI)", 2005.
[11] M. Renieris and S. Reiss, "Fault Localization with NearestNeighbor Queries", Proc. 18th IEEE Int'l Conf. Automated Software Eng. (ASE '03), 2003, pp. 30-39.
[12] P. Arumuga Nainar, T. Chen, J. Rosin, and B. Liblit, "Statistical debugging using compound Boolean predicates", In proceeding of International Symposium on Software Testing an Analysis, London, 2007.
[13] B. Liblit, A. Aiken, X. Zheng, and M.I. Jordan, "Bug isolation via remote program sampling", In Proceedings of the ACM SIGPLAN 2003 Conference on Programming Language Design and Implementation, San Diego, 2003, pp 141–154.
[14] S. J. Zeil,"Perturbation techniques for detecting domain errors", IEEE Transactions on Software Engineering, 15, June 1989, pp. 737-764.

[15] Bouchachia, A., Mittermeir, R., Siclecky, P., Stafiej, S. and Zieminski, M., "Nature-Inspired Techniques for Conformance Testing of Object-Oriented Software", Applied Soft Computing. J, 2009, pp. 1-16.
[16] Holzmann, J. G. and Smith, H. M., "A Practical Method for Verifying Event-Driven Software", 1998.

**Seyyede Roya Alavi** received the M.Sc degree in Computer Engineering in 2011 from Iran university of Islamic Azad University, Zanjan, Iran. She has published several papers about Software Testing. She is also lecturer at Azad University Khoy branch.

# Analysis of Quality of Service Performances of Connection Admission Control Mechanisms in OFDMA IEEE 802.16 Network using BMAP Queuing

**Abdelali EL BOUCHTI, Abdelkrim HAQIQ and Said EL KAFHALI**

**Computer, Networks, Mobility and Modeling laboratory**
**e- NGN research group, Africa and Middle East**
**FST, Hassan 1st University, Settat, Morocco**

**Abstract**

In this paper, we consider a single-cell IEEE 802.16 environment in which the base station allocates subchannels to the subscriber stations in its coverage area. The subchannels allocated to a subscriber station are shared by multiple connections at that subscriber station. To ensure the Quality of Service (QoS) performances, two Connection Admission Control (CAC) mechanisms, namely, threshold-based and queue-aware CAC mechanisms are considered at a subscriber station. A queuing analytical framework for these admission control mechanisms is presented considering Orthogonal Frequency Division Multiple Access (OFDMA) based transmission at the physical layer. Then, based on the queuing model, both the connection-level and the packet-level performances are studied and compared with their analogues in the case without CAC. The connection arrival is modeled by a Poisson process and the packet arrival for a connection by Batch Markov Arrival Process (BMAP). We determine analytically and numerically different QoS performance measures (connection blocking probability, average number of ongoing connections, average queue length, packet dropping probability, queue throughput and average packet delay).

***Keywords:*** IEEE 802.16/WiMAX, OFDMA, BMAP Process, Queuing Theory, Quality of Service Performances.

## 1. Introduction

Worldwide Interoperability for Microwave Access (WiMAX) ([1], [2]) is becoming one of the hottest topics in the development of wireless technology. Researchers and developers are focusing on the development of WiMAX base station technology which is expected to provide services in 2008 around the world. In the 4th quantum of 2005, the IEEE 802.16e specification was launched to the market detailing the full specification of mobile WiMAX. In essence, the Quality of Service (QoS) for WiMAX is desperately required. WiMAX aims at providing low latency, low delay/jitter, low loss, adequate bandwidth service. In general, satisfactory QoS always requires a high operational cost. It is known that both Connection Admission Control (CAC) and packet scheduling co-operate to provide a high QoS and a low-cost service. To ensure further the QoS of high priority services, packet scheduling grants the channel for service according to their priorities so that un-serviced packets will line-up in the buffer.

Orthogonal Frequency Division Multiple Access (OFDMA) [5] is a promising wireless access technology for the next generation broad-band packet networks. With OFDMA [27], which is based on orthogonal frequency division multiplexing (OFDM), the wireless access performance can be substantially improved by transmitting data via multiple parallel channels, and also it is robust to inter-symbol interference and frequency-selective fading. OFDMA has been adopted as the physical layer transmission technology for WiMAX [26] based broadband wireless networks. Although the WiMAX standard [4] defines the physical layer specifications and the Medium Access Control (MAC) signaling mechanisms, the radio resource management methods such as those for CAC and dynamic bandwidth adaptation are left open. However, to guarantee QoS performances (e.g., call blocking rate, packet loss, and delay), efficient admission control is necessary in a WiMAX network [25] at both the subscriber and the base stations.

To analyze various connection admission control algorithms, analytical models based on continuous-time Markov chain (CTMC) ([3], [6], and [9]), were proposed in [19]. However, most of these models dealt only with call/connection-level performances [28] for the traditional voice-oriented cellular networks. In addition to the connection-level performances, packet-level performances also need to be considered for data-oriented packet-switched wireless networks such as WiMAX networks.

An earlier relevant work was reported by the authors in [23], [12], and [13]. They considered a similar model in OFDMA based-WiMAX but they modeled packet-level by Poisson process and MMPP process ([14], [16], and [17]) and they compared various QoS measures of CAC mechanisms. Since the introduction of batch Markovian arrival process (BMAP) by Lucantoni [21] many authors investigated queuing models with BMAP ([10], [15]). The

reason is that BMAP enables more realistic and more accurate traffic modeling, since it can also capture dependency in traffic processes. Most of these works apply the standard matrix analytic-method pioneered by Neuts [18]. The incoming traffic has self similar and bursty nature also in wireless networks causing correlation in inter-arrival times, which influences the performance of the system. Our motivation for using BMAP is that it can model such traffic correlation. Hence applying BMAP in the queuing model enables the traffic correlation dependent performance evaluation of the system.

In this paper, we present two connection admission control mechanisms for a multi-channel and multi-user OFDMA network. The first mechanism is threshold-based, in which the concept of guard channel is used to limit the number of admitted connections to a certain threshold. The second mechanism, namely, queue-aware is based on the information on queue status and it also inherits the concept of fractional guard channel in which an arriving connection is admitted with certain connection acceptance probability. Specifically, the connection acceptance probability is determined based on the queue status (i.e., the number of packets in the queue). A queuing analytical model is developed based on a three- DTMC which captures the system dynamics in terms of the number of connections and queue status. We assume that the connection arrival and the packet arrival for a connection follow a Poisson process and a BMAP process respectively. Based on this model, various performance parameters such as connection blocking probability, average number of ongoing connections, average queue length, probability of packet dropping due to lack of buffer space, queue throughput, and average queuing delay are obtained. The numerical results reveal the comparative performance characteristics of the threshold-based and the queue-aware CAC and the without CAC algorithms in an OFDMA-based WiMAX network.

The remainder of this paper is organized as follows. Section 2 describes the system model including the objective of CAC policy. The formulation of the analytical model for connection admission control is presented in Section 3. In section 4 we determine analytically different performance parameters. Numerical results are stated in Section 5. Finally, section 6 concludes the paper.

## 2. Model description

### 2.1 System model

We consider a single cell in a IEEE 802.16/WiMAX network with a base station and multiple subscriber stations (Figure 1). Each subscriber station serves multiple connections. Admission control is used at each subscriber station to limit the number of ongoing connections through that subscriber station. At each subscriber station, traffic from all uplink connections are aggregated into a single queue [24]. The size of this queue is finite (i.e., $X$ packets) in which some packets will be dropped if the queue is full upon their arrivals. The OFDMA transmitter at the subscriber station retrieves the head of line packet(s) and transmits them to the base station. The base station may allocate different number of subchannels to different subscriber stations. For example, a subscriber station with higher priority could be allocated more number of subchannels.

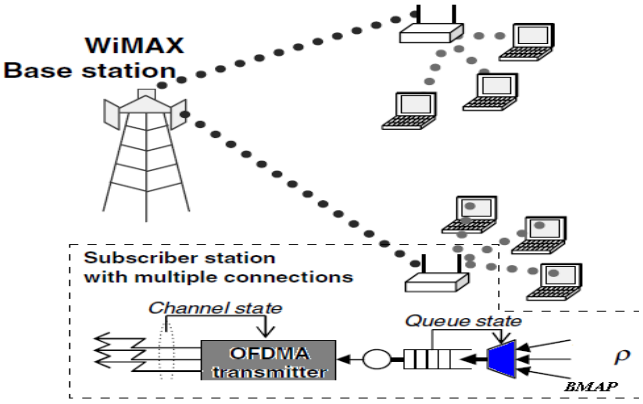



Figure 1: System model

### 2.1 CAC policy

The main objective of a CAC mechanism is to limit the number of ongoing connections/flows so that the QoS performances can be guaranteed for all the ongoing connections. Then, the admission control decision is made to accept or reject an incoming connection. To ensure the QoS performances of the ongoing connections, the following CAC mechanism for subscriber stations are proposed.

*2.1.1 Threshold-Based CAC mechanism*

In this case, a threshold C is used to limit the number of ongoing connections. When a new connection arrives, the CAC module checks whether the total number of connections including the incoming one is less than or equal to the threshold C. If it is true, then the new connection is accepted, otherwise it is rejected.

*2.1.2 Queue-Aware CAC mechanism*

This mechanism works based on connection acceptance probability $\alpha_x$ which is determined based on the queue status. Specifically, when a connection arrives, the CAC module accepts the connection with probability $\alpha_x$, where $x$ $(x \in \{0,1,...,X\})$ is the number of packets in the queue

in the current time slot. Here, *X* denotes the size of the queue of the subscriber station under consideration. Note that the value of the parameter $\alpha_x$ can be chosen based on the radio link level performance (e.g., packet delay, packet dropping probability) requirements.

## 3. Formulation of the Analytical Model

### 3.1 Formulation of the Queuing Model

An analytical model based on DTMC is presented to analyze the system performances at both the connection-level and at the packet-level for the connection admission ([7], [8], and [22]) mechanisms described before. We assume that packet arrival for a connection follows a BMAP process which is identical for all connections in the same queue. However, in the future we will consider non-homogenous Poisson process [29] as arrival process. The connection inter-arrival time and the duration of a connection are assumed to be exponentially distributed with average $1/\rho$ and $1/\mu$, respectively. In future, we will consider non exponential distributions using MRGP (Markov Re-Generative Process) [11] and/or phase-type expansions [29], [30].

The arrivals in the BMAP is directed by the irreducible continuous time Markov chain CTMC with a finite state space {0,1, …, *S*}. Sojourn time of the CTMC in the state *s* has exponential distribution with parameter $\lambda_s$. After time expires, with probability $p_0(s,s')$ the chain jumps into the state *s'* without generation of packets and with probability $p_k(s,s')$ the chain jumps into the state *s'* and a batch consisting of k packets is generated, $k \geq 1$. The introduced probabilities satisfy conditions: $p_0(s,s) = 0$, the sum of the probabilities of all outgoing transitions has to be equal to 1,

$$\sum_{k=1}^{\infty}\sum_{s'=0}^{S} p_k(s,s') + \sum_{\substack{s'=0 \\ s'\neq s}}^{S} p_0(s,s') = 1,\ 0 \leq s \leq S. \quad (1)$$

The infinitesimal generator of BMAP is given as

$$Q_{BMAP} = \begin{pmatrix} \mathrm{D}_0 & \mathrm{D}_1 & \mathrm{D}_2 & \mathrm{D}_3 & \cdots \\ 0 & \mathrm{D}_0 & \mathrm{D}_1 & \mathrm{D}_2 & \cdots \\ 0 & 0 & \mathrm{D}_0 & \mathrm{D}_1 & \cdots \\ 0 & 0 & 0 & \mathrm{D}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (2)$$

Where the matrices $D_k$ are given by

$$\mathrm{D}_0 = [D_{ss'}],\ 0 \leq s \leq \mathrm{S},\ 0 \leq s' \leq \mathrm{S} \quad (3)$$

$$\mathrm{D}_k = [D_{k,ss'}],\ 0 \leq s \leq \mathrm{S},\ 0 \leq s' \leq \mathrm{S},\ k \geq 0 \quad (4)$$

Where:

$$D_{k,ss'} = \lambda_s p_k(s,s'),\ \ 0 \leq s \leq \mathrm{S},\ 0 \leq s' \leq \mathrm{S},\ \ k > 0. \quad (5)$$

$$D_{ss} = -\lambda_s, \qquad D_{ss'} = \lambda_s p(s,s'),\ \ s \neq s' \quad (6)$$

Knowing the matrices $\mathrm{D}_k$ the infinitesimal generator D can be defined as

$$\mathrm{D} = \sum_{k=0}^{\infty} \mathrm{D}_k. \quad (7)$$

The steady-state probability vector $\pi_{BMAP}$ of the CTMC with generator D can be calculated as usual:

$$\pi_{BMAP}.\mathrm{D} = 0, \quad \pi_{BMAP}.e = 1. \quad (8)$$

Here and below *e* is the column vector of appropriate dimension consisting of all 1's.
The mean steady-state arrival rate generated by the BMAP is:

$$\lambda_{BMAP} = \pi_{BMAP} \sum_{k=1}^{\infty} k\mathrm{D}_k e. \quad (9)$$

The state of the system is described by the process $Y_t = (S_t, X_t, C_t)$, where $S_t$ is the state (phase) of an irreducible continuous time Markov chain, $X_t$ is the number of packets in the aggregated queue and $C_t$ the number of ongoing connections at the end of every time slot *t*.

Thus, the state space of the system for both the CAC mechanisms is given by:

$$E = \{(s,x,c) / s \in \{1,....,S\},\ 0 \leq x \leq X,\ \mathrm{c} \geq 0\}\ .$$

For the both CAC algorithms, the number of packet arrivals depends on the number of connections. However, for the queue-aware CAC algorithm, the number of packets in the queue affects the acceptance probability for a new connection. The state transition diagram is shown in figure 2.

Note that the probability that *n* Poisson events with average rate $\rho$ occur during an interval T can be obtained as follows:

$$f_n(\rho) = \frac{e^{-\rho T}(\rho T)^n}{n!} \quad (10)$$

This function is required to determine the probability of both connection and packet arrivals.

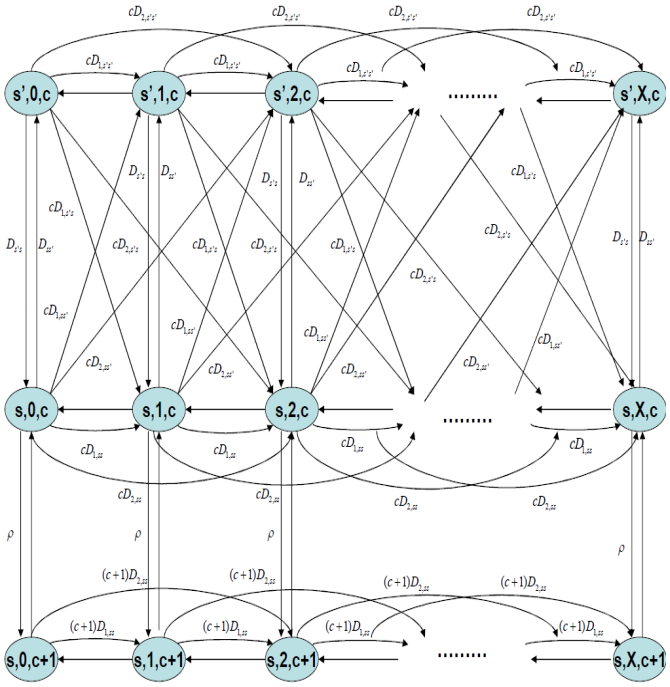



Figure 2: State transition diagram of discrete time Markov chain.

## 3.2 Threshold-Based CAC Algorithm

In this case, the transition matrix $Q$ for the number of connections in the system can be expressed as follows:

$$Q = \begin{bmatrix} q_{0,0} & q_{0,1} & & & & \\ q_{0,1} & q_{1,1} & q_{1,2} & & & \\ \ddots & \ddots & \ddots & & & \\ & & q_{C-2,C-1} & q_{C-1,C-1} & q_{C-1,C} \\ & & & q_{C-1,C} & q_{C,C} \end{bmatrix} \quad (11)$$

where each row indicates the number of ongoing connections. As the length of a frame $T$ is very small compared with connection arrival and departure rates, we assume that the maximum number of arriving and departing connections in a frame is one. Therefore, the elements of this matrix can be obtained as follows:

$$\begin{aligned} q_{c,c+1} &= f_1(\rho)\times(1-f_1(c\mu)), \quad c=0,1,...,C\text{-}1 \\ q_{c,c-1} &= (1-f_1(\rho))\times f_1(c\mu), \quad c=1,2,...,C \\ q_{c,c} &= f_1(\rho)\times f_1(c\mu)+(1-f_1(\rho))\times(1-f_1(c\mu)), \quad c=0,1,...,C \end{aligned} \quad (12)$$

where $q_{c,c+1}$, $q_{c,c-1}$ and $q_{c,c}$ represent the cases that the number of ongoing connections increases by one, decreases by one, and does not change, respectively.

## 3.3 Queue-Aware CAC Algorithm

Because the admission of a connection in this case depends on the current number of packets in the queue, the transition matrix can be expressed based on the number of packets ($x$) in the queue as follows:

$$Q_x = \begin{bmatrix} q^{(x)}_{0,0} & q^{(x)}_{0,1} & & \\ q^{(x)}_{1,0} & q^{(x)}_{1,1} & q^{(x)}_{1,2} & \\ q^{(x)}_{2,0} & q^{(x)}_{2,1} & q^{(x)}_{2,2} & \\ & \ddots & \ddots & \ddots \end{bmatrix} \quad (13)$$

where:

$$\begin{aligned} q^{(x)}_{c,c+1} &= f_1(\alpha_x\rho)\times(1-f_1(c\mu)), \quad c=0,1,... \\ q^{(x)}_{c,c-1} &= (1-f_1(\alpha_x\rho))\times f_1(c\mu), \quad c=1,2,... \\ q^{(x)}_{c,c} &= f_1(\alpha_x\rho)\times f_1(c\mu)+(1-f_1(\alpha_x\rho))\times(1-f_1(c\mu)), \quad c=0,1,... \end{aligned} \quad (14)$$

in which $\alpha_x$ is the connection acceptance probability when there are x packets in the queue.

## 3.4 Transition Matrix for the Queue

The transition matrix $P$ of the entire system can be expressed as in equation 15. The rows of matrix $P$ represent the number of packets ($x$) in the queue.

$$P = \begin{bmatrix} p_{0,0} & \cdots & p_{0,A} & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ p_{R,0} & \cdots & p_{R,R} & \cdots & p_{R,R+A} & \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & p_{x,x-R} & \cdots & p_{x,x} & \cdots & p_{x,x+R} \\ & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \quad (15)$$

Matrices $p_{x,x'}$ represent the changes in the number of packets in the queue (i.e., the number of packets in the queue changing from $x$ in the current frame to $x'$ in the next frame). We first establish matrices $\mathrm{v}_{(s,x),(s,x')}$, where the diagonal elements of these matrices are given in equation 16. For $r \in \{0,1,2,...,L\}$ and $n \in \{0,1,2,...,(c\times A)\}$, $l=1,2,...,L$, and $m=1,2,...,(c\times A)$. The non-diagonal elements of $\mathrm{v}_{(s,x),(s,x')}$ are all zero.

$$\begin{aligned} \left[\mathrm{v}_{(s,x);(s,x-l)}\right]_{c,c} &= \sum_{n-r=l}\sum_{k=1}^{\infty}\sum_{s'=0}^{S} f_n((c-1)D_{k,ss'})[R]_r \\ \left[\mathrm{v}_{(s,x);(s,x+m)}\right]_{c,c} &= \sum_{r-n=m}\sum_{k=1}^{\infty}\sum_{s'=0}^{S} f_n((c-1)D_{k,ss'})[R]_r \\ \left[\mathrm{v}_{(s,x);(s,x)}\right]_{c,c} &= \sum_{r=n}\sum_{k=1}^{\infty}\sum_{s'=0}^{S} f_n((c-1)D_{k,ss'})[R]_r \end{aligned} \quad (16)$$

Note that, matrix **R** [15] has size $1 \times R + 1$, where $R$ indicates the maximum number of packets that can be transmitted in one frame. Here, $A$ is the maximum number of packets that can arrive from one connection in one frame and $L$ is the maximum number of packets that can be transmitted in one frame by all of the allocated sub channels allocated to that particular queue and it can be obtained from $L = \min(R, x)$. This is due to the fact that the maximum number of transmitted packets depends on the number of packets in the queue and the maximum possible number of transmissions in one frame. Note that, $\left[\mathrm{v}_{(s,x);(s,x-l)}\right]_{c,c}$, $\left[\mathrm{v}_{(s,x);(s,x+m)}\right]_{c,c}$ and $\left[\mathrm{v}_{(s,x);(s,x)}\right]_{c,c}$ represent the probability that the number of packets in the queue increases by $l$, decreases by $m$, and does not change, respectively, when there are $c-1$ ongoing connections. Here, $[\mathrm{v}]_{i,j}$ denotes the element at row $i$ and column $j$ of matrix v, and these elements are obtained based on the assumption that the packet arrivals for the ongoing connections are independent of each other.

Finally, we obtain the matrices $p_{x,x'}$ by combining both the connection-level and the queue-level transitions as follows:

$$p_{x,x'} = Q\mathrm{v}_{(s,x),(s,x')} \quad (17)$$

$$p_{x,x'} = Q_x\mathrm{v}_{(s,x),(s,x')} \quad (18)$$

for the cases of threshold-based (Equation 11) and queue-aware (Equation 13) CAC algorithms, respectively.

## 4. Performance Parameters

In this section, we determine the connection-level and the packet-level performance parameters (i.e., connection blocking probability, average number of ongoing connections in the system, and average queue length) for the both CAC mechanisms.

For the threshold-based CAC mechanism, all of the above performance parameters can be derived from the steady state probability vector of the system states $\pi$, which is obtained by solving $\pi P = \pi$ and $\pi 1 = 1$, where $1$ is a column matrix of ones. However, for the queue-aware CAC algorithm, the size of the matrix $Q_x$ needs to be truncated at $C_{tr}$ (i.e., the maximum number of ongoing connections at the subscriber station).

Also, the size of the matrix $P$ needs to be truncated at $X$ (i.e., the maximum number of packets in the queue) for the both mechanisms.

The steady-state probability, denoted by $\pi(s, x, c)$ for the state that there are $c$ connections and $x \in \{0,1,..., X\}$ packets in the queue, can be extracted from matrix $\pi$ as follows

$$\pi(s,x,c) = [\pi]_{s\times x\times((C+1)+c)}, \quad s = 1,...,S;\ c = 0,1,...,C' \quad (19)$$

where $C' = C$ and $C' = C_{tr}$ for the threshold-based and the queue-aware CAC algorithms, respectively. Using these steady state probabilities, the various performance parameters can be obtained. Note that, the subscripts *tb* and *qa* are used to indicate the performance parameters for the threshold-based and the queue-aware CAC mechanisms, respectively.

### 4.1 Connection Blocking Probability

This performance parameter indicates that an arriving connection will be blocked due to the admission control decision. It indicates the accessibility of the wireless service, and for the threshold-based CAC mechanism. It can be obtained as follows:

$$p^{tb}_{block} = \sum_{s=1}^{S}\sum_{x=0}^{X}\pi(s, x, C). \quad (20)$$

The above probability refers to the probability that the system serves the maximum allowable number of ongoing connections.

The blocking probability for the queue-aware CAC mechanism is obtained from

$$p^{qa}_{block} = \sum_{s=1}^{S}\sum_{x=0}^{X}\sum_{c=1}^{C_{tr}}((1-\alpha_x).\pi(s, x, C)). \quad (21)$$

in which the blocking probability is the sum of the probabilities of rejection for all possible number of packets in the queue.

### 4.2 Average Number of Ongoing Connections

It can be obtained as

$$N^{tb}_{c} = \sum_{s=1}^{S}\sum_{x=0}^{X}\sum_{c=0}^{C} c.\pi(s, x, c) \quad (22)$$

$$N^{qa}_{c} = \sum_{s=1}^{S}\sum_{x=0}^{X}\sum_{c=0}^{C_{tr}} c.\pi(s, x, c) \quad (23)$$

### 4.3 Average Queue Length Average

It is given by

$$N_x^{tb} = \sum_{s=1}^{S} \sum_{x=0}^{C} \sum_{x=0}^{X} x.\pi(s,x,c) \quad (24)$$

$$N_x^{qa} = \sum_{s=1}^{S} \sum_{x=0}^{C_{tr}} \sum_{x=0}^{X} x.\pi(s,x,c) \quad (25)$$

### 4.4 Packet Dropping Probability

This performance parameter refers to the probability that an incoming packet will be dropped due to the unavailability of buffer space. It can be derived from the average number of dropped packets per frame. Given that there are *x* packets in the queue and the number of packets in the queue increases by *m*, the number of dropped packets is $m-(X-x)$ for $m > X-x$ , and zero otherwise. The average number of dropped packets per frame is obtained as follows:

$$N_{drop} = \sum_{s=0}^{S}\sum_{c=1}^{C}\sum_{x=0}^{X}\sum_{m=X-x+1}^{A}\left(\sum_{l=1}^{C}[p_{x,x+m}]_{c,l}\right).(m-(X-x)).\pi(s,x,c) \quad (26)$$

where the term $\left(\sum_{l=1}^{C}[p_{x,x+m}]_{c,l}\right)$ indicates the total probability that the number of packets in the queue increases by *m* at every arrival phase. Note that, we consider probability $p_{x,x+m}$ rather than the probability of packet arrival as we have to consider the packet transmission in the same frame as well.

After calculating the average number of dropped packets per frame, we can obtain the probability that an incoming packet is dropped as follows:

$$p_{drop} = \frac{N_{drop}}{\overline{\lambda}} \quad (27)$$

where $\overline{\lambda}$ is the average number of packet arrivals per frame and it can be obtained from

$$\overline{\lambda} = \lambda_{BMAP} N_c \quad (28)$$

### 4.5 Queue throughput

It measures the number of packets transmitted in one frame and can be obtained from

$$\varphi = \lambda_{BMAP}(1 - p_{drop}). \quad (29)$$

### 4.6 Average Packet Delay

It is defined as the number of frames that a packet waits in the queue since its arrival before it is transmitted. We use Little's law [22] to obtain average delay as follows:

$$D = \frac{N_x}{\varphi} \quad (30)$$

## 5. Numerical Results

In this section we present the numerical results of both CAC mechanisms. We use the Matlab software to solve numerically and to evaluate the various performance parameters.

### 5.1 Parameter Setting

We consider one queue (which corresponds to a particular subscriber station) for which five sub-channels are allocated and we assume that the average SNR is the same for all of these sub-channels. Each sub-channel has a bandwidth of 160 kHz. The length of a subframe for downlink transmission is one millisecond, and therefore, the transmission rate in one subchannel with rate ID = 0 (i.e., BPSK modulation and coding rate is 1/2) is 80 kbps. We assume that the maximum number of packets arriving in one frame for a connection is limited to 50.

For the threshold-based CAC mechanism, the value of the threshold C is varied according to the evaluation scenarios. For the queue-aware CAC mechanism, the value of the connection acceptance probability is determined as follows:

$$\alpha_{x=} \begin{cases} 1, & 0 \le x \prec B_{th} \\ 0, & B_{th} \le x \le X. \end{cases} \quad (30)$$

In the performance evaluation, we use $B_{th} = 100$.

For performance comparison, we also evaluate the queuing performance in the absence of CAC mechanism. For the case without CAC, we truncate the maximum number of ongoing connections at 70. The average duration of a connection is set to twenty minutes for all the evaluation scenarios. The queue size is 300 packets. The parameters are set as follows: The connection arrival rate is 0.9 connections per minute the batch packets of size 30 (i.e., *k*=30). Average SNR on each sub-channel is 5 db.
For clarity, the all numerical parameters are summarized in table 1.

Table 1: Summary of numerical parameters.

| | |
|---|---|
| $X$ | 300 |
| $A$ | 50 |
| $C_{tr}$ | 70 |
| $B_{th}$ | 100 |
| $\rho$ | 0.9 |
| $k$ | 30 |
| $\mu$ | 20 |
| Average SNR | 5 dB |

Note that, we vary some of these parameters depending on the evaluation scenarios whereas the others remain fixed.

## 5.2 Performance of CAC policy

We first examine the impact of connection arrival rate on connection-level performances. Variations in average number of ongoing connections and connection blocking probability with connection arrival rate are shown in Figures 3 and 4. As expected, when the connection arrival rate increases, the number of ongoing connections and connection blocking probability increase.

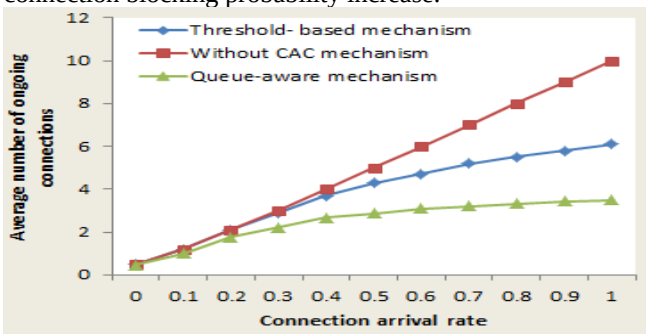



Figure 3: Average number of ongoing connections under different connection arrival rates.

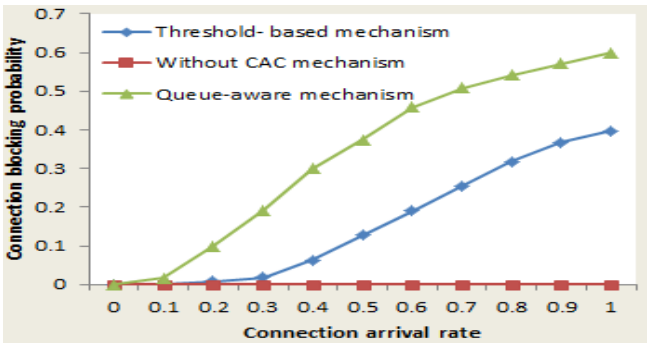



Figure 4: Connection blocking under different connection arrival rates.

The packet-level performances under different connection arrival rates are shown in Figures 5 through 8 for average number of packets in the queue, packet dropping probability, queue throughput, and average queuing delay, respectively. These performance parameters are significantly impacted by the connection arrival rate. Because the both CAC mechanisms limit the number of ongoing connections, packet-level performances can be maintained at the target level. In this case, both CAC mechanisms result in better packet-level performances compared with those without CAC mechanism.

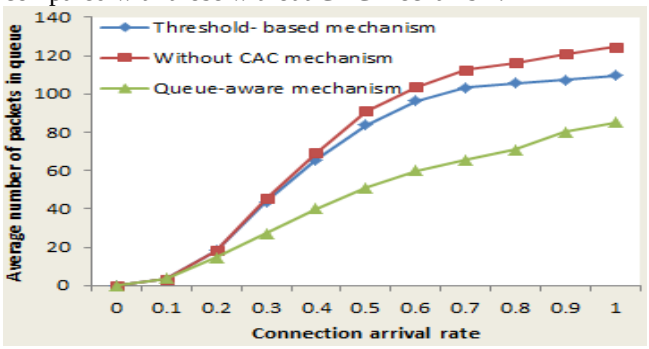



Figure 5: Average number of packets in queue under different connection rates.

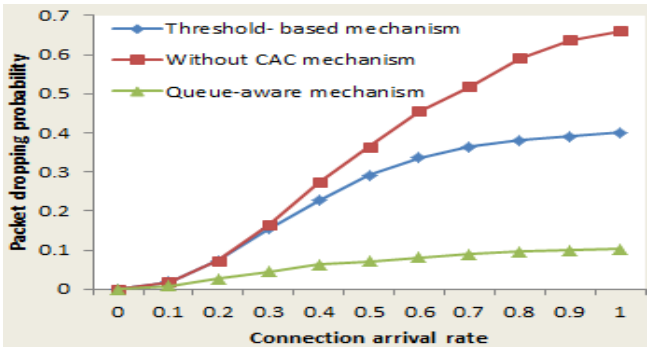



Figure 6: Packet dropping under different connection arrival rates.

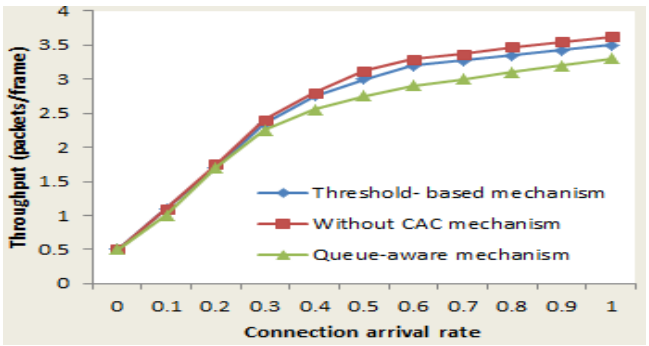



Figure 7: Queuing throughput under different connection arrival rates.

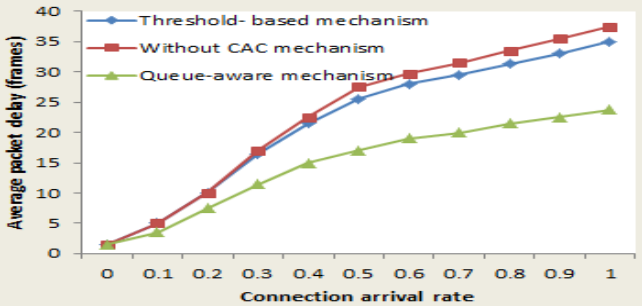



Figure 8: Average packet delay under different connection arrival rates.

Variations in packet dropping probability and average packet delay with channel quality are shown in Figures 9 and 10. As expected, the packet-level performances become better when channel quality becomes better.

Also, we observe that the connection-level performances for the threshold-based CAC mechanism and those without CAC mechanism are not impacted by the channel quality when this later becomes better (the connection blocking probability remains constant when the channel quality varies), connection blocking probability decreases significantly for the queue-aware CAC mechanism when the channel quality becomes better (Figure. 11).

Based on these observations, we can conclude that the queue-aware CAC can adapt the admission control decision based on the queue status which is desirable for a system with high traffic fluctuations.

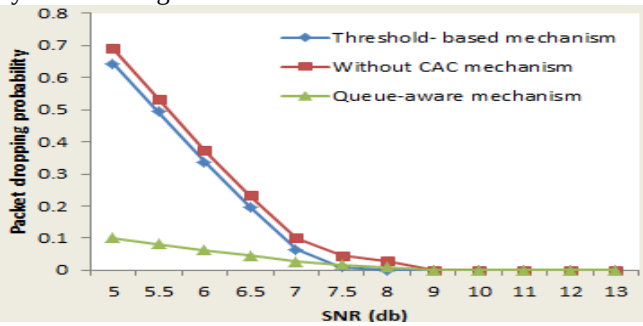



Figure 9: Packet dropping probability under different channel qualities.

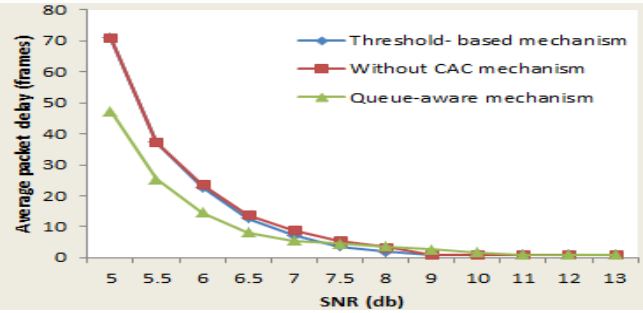



Figure 10: Average packet delay under different channel qualities.

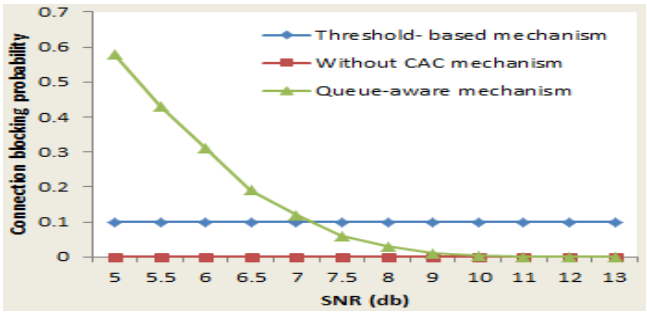



Figure 11: Connection blocking probability under different channel qualities.

## 6. Conclusion

In this paper, we have addressed the problem of queuing theoretic performance modeling and analysis of OFDMA transmission under admission control. We have considered a IEEE 802.16/WiMAX system model in which a base station serves multiple subscriber stations and each of the subscriber stations is allocated with a certain number of subchannels by the base station. There are multiple ongoing connections at each subscriber station.

We have presented two connection admission control mechanisms for a multi-channel and multi-user OFDMA network, namely, queue-aware mechanism and threshold-based mechanism. While the threshold-based CAC mechanism simply fixes the number of ongoing connections, the queue-aware CAC mechanism considers the number of packets in the queue for the admission control decision of a new connection. The connection-level and packet-level performances of these CAC mechanisms have been studied based on the queuing model.

The connection-level and packet-level performances of the both CAC mechanisms have been studied based on the queuing model. The connection arrival is modeled by a Poisson process and the packet arrival for a connection by a BMAP process. We have determined analytically and numerically different performance parameters, such as connection blocking probability, average number of ongoing connections, average queue length, packet dropping probability, queue throughput, and average packet delay.

Numerical results show that, the performance parameters of connection-level and packet-level are significantly impacted by the connection-level rate, the both CAC mechanisms results in better packet-level performances compared with those without CAC mechanism. The packet-level performances become better when channel quality becomes better. On the other hand,

the connection-level performances for the threshold-based CAC mechanism and those without CAC mechanism are not impacted by the channel quality when this later becomes better. Then, the queue-aware CAC can adapt the admission control decision based on the queue status which is desirable for a system with high traffic fluctuations.

All the results showed in this paper remain in correlation with those presented in [23], [12] and [13] even if we change here the arrival packet Poisson process by an MMPP process or by BMAP process, which is more realistic.

## References

[1] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air

Interface for Fixed Broadband Wireless Access Systems, IEEE 802.16- 2004, 2004.

[2] IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005, IEEE Standard for Local and metropolitan area networks-Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Dec. 7, 2005.

[3] K. Astrom "Introduction to Stochastic Control Theory," Dover Publications: New York, 2006.

[4] B. Baynat, S. Doirieux, G. Nogueira, M. Maqbool, and M. Coupechoux, "An efficient analytical model for wimax networks with multiple traffic profiles," in Proc. of ACM/IET/ICST IWPAWN, September 2008.

[5] B. Baynat, G. Nogueira, M. Maqbool, and M. Coupechoux, "An efficient analytical model for the dimensioning of wimax networks," in Proc. of 8th IFIP-TC6 Networking Conference, May 2009.

[6] L. B. Le, E. Hossain, and A. S. Alfa, "Queuing analysis for radio link level scheduling in a multi-rate TDMA wireless network," in Proc. IEEE GLOBECOM'04, vol. 6, pp. 4061–4065, November–December 2004.

[7] D. Bertsekas Dynamic Programming and Optimal Control. Athena Scientific: Belmont, MA, U.S.A., 2005.

[8] J. Birge, F. Louveaux Introduction to Stochastic Programming. Springer: Berlin, 1997.

[9] L. B. Le, E. Hossain, and A. S. Alfa, "Queuing analysis for radio link level scheduling in a multi-rate TDMA wireless network," in Proc. IEEE GLOBECOM'04, vol. 6, pp. 4061–4065, November–December 2004.

[10] G. Bolch, S. Greiner, H. de meer, and K.S. Trivedi, " Qeueing network and Markov chain, Second edition. John Wiley, 2006.

[11] H, Choi, V. G. Kulkarni, and K. S. Trivedi, " Markov regenerative stochastic petri nets" Performance Evaluation, 1994.

[12] Abdelali EL BOUCHTI, Said EL KAFHALI, and Abdelkrim HAQIQ "Performance Analysis of Connection Admission Control Scheme in IEEE 802.16 OFDMA Networks" IJCSIS, Vol. 9, No. 3, pp. 45-51 March 2011.

[13] Abdelali EL BOUCHTI, Said EL KAFHALI, and Abdelkrim HAQIQ "Performance Modeling and Analysis of Connection Admission Control in OFDMA based WiMAX System with MMPP Queueing" WCSIT, Vol. 1, No. 4, pp. 148-156 , 2011.

[14] Abdelali EL BOUCHTI and Abdelkrim HAQIQ, " Comparaison of two Access Mechanisms for Multimedia Flow in High Speed Downlink Packet Access Channel", IJAET, Vol 4, Issue 2, pp. 29-35, 2011.

[15] Abdelali EL BOUCHTI and Abdelkrim HAQIQ, "Access Control of Mulimedia traffic in a 3.5 G Wireless Network" , proceedings of ICMCS'11, Ouarzazate, April 7-9 2011.

[16] Abdelali EL BOUCHTI and Abdelkrim Haqiq, "The performance evaluation of an access control of heterogeneous flows in a channel HSDPA", proceedings of CIRO'10, Marrakesh, Morocco, 24-27 May 2010.

[17] Abdelali EL BOUCHTI, Abdelkrim Haqiq, Mohamed Hanini and Mohamed Elkamili "Access Control and Modeling of Heterogeneous Flow in 3.5G Mobile Network by using MMPP and Poisson processes", proceedings of MICS'10, Rabat, Morocco, 2-4 November 2010.

[18] M. F. Neuts. \Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach," The John Hopkins University Press, Baltimore, 1981.

[19] Y. Fang and Y. Zhang, "Call admission control schemes and performance analysis in wireless mobile networks," IEEE Transactions on Vehicular Technology, vol. 51, no. 2, March 2002, pp. 371–382.

[20] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," IEEE Transactions on Vehicular Technology, pp. 77–92, August 1986.

[21] D. L. Lucantoni. New results on the single server queue with a batch markovian arrival process, Stochastic Models, 7 (1991), 1-46.

[22] R. Nelson, "Probability, stochastic process, and queueing theory", Springer-Verlag, third printing, 2000.

[23] D. Niyato and E. Hossain, "Connection admission control in OFDMA-based WiMAX networks: Performance modeling and analysis," invited chapter in WiMax/MobileFi: Advanced Research and Technology, (Ed. Y. Xiao), Auerbach Publications, CRC Press, December 2007.

[24] D. Niyato and E. Hossain, "Connection admission control algorithms for OFDMA wireless networks," in Proc. IEEE GLOBECOM'05, St. Louis, MO, USA, 28 November–2 December 2005.

[25] H.Okamura, T. Dohi, and K.S. Trivedi, , "Markovian arrival process parameter estimation with group data," IEEE/ACM Transactions on Networking, vol. 17, no. 4, pp. 1326-1339, August 2008.

[26] D. Pareek, "WiMax: Taking Wireless to the MAX," Auerbach Publishers Inc. June, 2006.

[27] P. Piggin, "Wimax in depth: Broadband wireless access," IEEE Commun. Eng., vol. 2, no. 5, pp. 36–39, Oct. 2004.

[28] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," in Proc. IEEE INFOCOM'96, vol. 1, San Francisco, CA, March 1996, pp. 43–50.

[29] K. S. Trivedi, Probability and statistics with reliability, Queuing and Computer Science Applications, second edition. Wiley, 2001.

[30] D. Wang, R. M. Fricks, and K. S. Trivedi, "Dealing with non-exponential distributions in dependability models," in Symp. On Performance Evaluation- Stories and Perspectives. G. Kotsis (ed.), Oesterreichchische Computer Gessellschaft, 2003, pp. 273-302.

# Modelling Efficient Process Oriented Architecture for Secure Mobile Commerce Using Hybrid Routing Protocol in Mobile Adhoc Network

Chitra Kiran N.
*Research Scholar*
, Dept. of Electronics & Communication Engg ,
*UVCE*
Bangalore, India

Dr. G. Narendra Kumar
*Prof.* Dept. of Electronics & Communication Engg.
UVCE
Bangalore, India

***Abstract*— The proposed research work presents a novel approach of process oriented architecture for secure mobile commerce framework using uniquely designed hybrid mobile adhoc routing protocols using reactive and proactive type in real time test-bed. The research work discusses about deployment of mobile commerce which is one of the emerging trend in mobile applications with huge demands. Majority of the existing system lacks either QoS or efficient security protocol when it relates to secure mobile transaction due to the reason that development in wireless technology involved in m-commerce is still in its nascent stage. The real time test bed has been implemented with 20 Intel Atom processor with 32 bit OS establishing an adhoc network and by providing a random mobility to achieve any file type transfer from node to node. For the real-time set up purpose, the experiment is conducted in wireless infrastructure with mobility using G-based Linksys wireless router. Iteration of experiments conducted shows a satisfactory results. This research journal will provide insights with various parameters, security requirements, and concepts which is required in creating a robust model for secure m-commerce system.**

***Keywords-m-Commerce, Mobile adhoc network, AODV, DSDV, online transaction***

## I. INTRODUCTION

The current research concern is to design an ideal application to facilitate and utilize mobile commerce on the go and mobile adhoc network can be the most preferred platform to design such application. But to best of our knowledge, the feasibility is still not confirmed in any prior research. Mobile adhoc network (MANET) is a network consisting of a collection of nodes capable of communicating with each other without any assistance from a network infrastructure. Applications of MANETs include the battlefield applications, rescue work, as well as civilian applications like an outdoor meeting, or an ad-hoc classroom. Several medium access control (MAC) protocols are used in wireless networks to manage the use of the wireless medium. M-commerce transactions can be performed over mobile adhoc networks or adhoc m-commerce which can be considered as wireless trading outside established computer networks [1]. Adhoc network is in constant use in the field of cost-effective communication. The utility of the adhoc wireless network is found to be effective enablers for engaging the clients in mobile commerce secure operations irrespective of time and location by making use of the wireless resources without any dependency for a network service provider. Along with the advantages, adhoc network also comes with lots of implementation challenges in its application. Recently, adhoc network has also been explored for mobile commerce application for easy communication system. Unfortunately, such types of networks are easily prone for serious network attacks, which make the client's private resources vulnerable. Therefore there is a huge requirement of guidelines for designing an analytical system which should assist to solve all the critical issues. The security issue which might be surfaced in the design of secure mobile commerce application will be inappropriate design of routing protocols in mobile adhoc network due to its dynamic topology and faster consumption of power in the mobile nodes which will render malicious activity to the client's device. Examples of such protocol will include the Bluetooth MAC layer protocol [2] and IEEE 802.11MAC layer protocol [3]. Because radio range is usually limited and the network components may have some mobility, the topology of a wireless network can vary with time. According to the relative mobility of hosts and routers, there are three different types of wireless networks e.g. Fixed wireless network [4], Wireless network with fixed access points [5], Mobile adhoc network [6]. Generally speaking, conventional routing protocols that are deployed in wired networks can only maintain routing in fixed wireless networks and mobile networks with fixed access points. Only one-hop routing is required over a link in a wireless network with fixed access points and many fixed wireless network [7]. The process of routing in mobile adhoc networks and certain fixed wireless networks exercise multiple-hop routing majorly [8]. Therefore the routing protocols for such types kind of wireless network should be able to maintain paths to other nodes and in

maximum cases, must tackle changes in paths due to mobility reasons. Conventional routing will not properly support routing in a MANET. Traditional routing protocols, such as Open Shortest Path First (OSPF) [9], can be deployed in some wireless networks with infrastructure. However, they cannot be directly employed to majority of MANETs since they are completely based on certain assumptions that will be only applicable in wired networks. A routing protocol may need to balance traffic based on the traffic load on links [10]. Scalability of routing protocols is an important issue for large networks [11]. The routing protocol may need to implement security to protect against attacks, such as sniffer, man-in-the-middle, or denial of service [12]. Researchers are designing new MANET routing protocols and comparing and improving existing MANET routing protocols before any routing protocols are standardized using simulations. The partial taxonomy of routing protocols in MANET is as shown below:

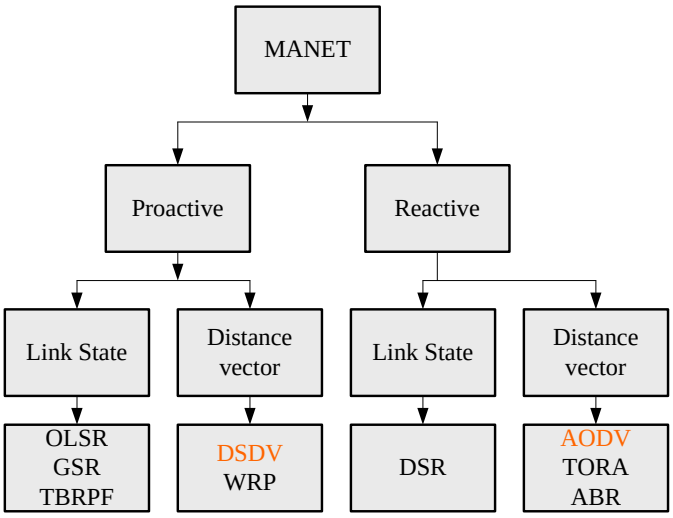



Fig.1. Partial Taxonomy of routing protocols in MANET

Also, it has been seen that the simulation results from different research groups are not consistent with each other [12]. This is because of a lack of consistency in MANET routing protocol models and application environments, including networking and user traffic profiles. Therefore results obtained by conducting experiments on routing protocol in simulations vary a lot in comparison to results obtained in real time scenario. Thus, the simulation scenarios will not be reasonable for all set of protocols and research implications cannot be generalized. Furthermore, it is most complicated for one to choose a proper routing protocol for a defined MANET application. The routing protocol is therefore desired to possess following individuality in order to be efficient: distributed operation, loop-freedom, demand-based operation, proactive operation, security, "sleep" period operation, unidirectional link support. [13].

Owing to the issues in difference in results in simulation and in real time approach which is a very challenging one, this research journal will highlight a novel approach of creating a real time design of process oriented architecture of mobile commerce application for mobility aware adhoc network using hybrid protocols. This process oriented architecture proposed in this research journal will also endeavor to highlight various security issues in mobile adhoc wireless mobile commerce system which will help the researchers for distinguishing various obstructions for solving such security issues in mobile commerce application. The main contribution of the research journal will be to create a categorization of various parameters identifying mobile commerce in mobile adhoc wireless network. All the operational parameters in mobile adhoc wireless mobile commerce will be identified. Any critical problems which may restrict the development of secure adhoc wireless mobile commerce system will also be focused.

In Section II, we will discuss about the previous research work in this area followed by Section III about mobile commerce system over wireless network. Section IV highlights about the feasible application for the proposed design work key issues followed by Problem Discussion in section V. In depth discussion of proposed system is done in section-VI followed by experimental set up in section-VII. Section VIII highlights the performance analysis of the conducted experiment followed by security requirement analyzation in section-IX and conclusion in section-X

## II. RELATED WORK

Routing protocols for diverse types of wireless networks have been discussed by a number of researchers. [12] The literature related to routing schemes used in MANET is classified into three main categories of routing protocols-Proactive (table-driven) protocols, Reactive (source-initiated) protocols, and Hybrid protocols. Proactive routing protocols tend to keep an up-to-date topological map of the entire network. With this map, the route is known and immediately available when a packet needs to be sent. The approach is equivalent to the one used in wired IP networks. Example of proactive routing protocols are Destination-Sequenced Distance-Vector routing (DSDV) [14] protocol, Cluster head Gateway Switch Routing (CGSR), Optimized Link-State Routing Protocol (OLSR) [15], Wireless Routing protocol (WRP) [16]. In distinction to proactive routing, reactive routing does not attempt to continuously determine the network connectivity. Instead, a route determination procedure is invoked on demand when a packet needs to be forwarded. The technique relies on queries that are flooded throughout the network. Reactive route determination is used in the Temporally Ordered Routing Algorithm (TORA), the Dynamic Source Routing (DSR) [17] and the Ad-hoc On-demand Distance Vector (AODV) [18] protocols. Another type of routing scheme is Hybrid routing

which combines the best features of both proactive and reactive approaches. Examples of such kind of protocols are Zone Routing Protocol (ZRP) [19],[20], Distributed Dynamic Routing algorithm (DDR). The Zone Routing Protocol, as its name implies, is based on the concept of zones. A routing zone is defined for each node separately, and the zones of neighboring nodes overlap. The routing zone has a radius r expressed in hops. Cheng e.t. al [21] has highlighted the vulnerability of the famous AODV routing protocol in his recent work. Rakesh [22] has presented and examined analytical simulation results for the routing protocols DSR and TORA network performance, using the well known network simulator OPNET 10. 0. Shukla [23] presents a novel method to enhance route maintenance part of DSR protocol. Javad [24] has proposed an algorithm for multicast routing protocol in wireless adhoc network using learning automata. Rafael [25] presents a performance analysis of different mobile payment protocols. The performance analysis includes the computational cost required by each entity to perform all the cryptographic operations and the transmission time required to transmit each message. But the work did not considered analysis using mobile payment protocols using elliptical curve cryptography. Suresh Chari e.t. al. [26] have identified some frameworks and their inherent exposures in security issues in m-commerce. Alia Fourati [27] has worked on secure and fair auction over adhoc network. Even in this work also, some specific security issues to adhoc networks were not treated. Osman [28] presents a fully distributed and self-organizing approach to managing group membership in such a loose trading community. It is designed to suit the dynamic nature of ad hoc wireless networking and the social characteristics of ad hoc m-commerce.

Obviously it can be seen in this section of related work that majority of research work from old to recent has been performed either in programming simulation or in commercial simulators like NS-2, OMNet, or OPNET. Almost no reported significant research work being carried out in routing protocols enhancements in real-time test-bed. This research journal will focus on implementation and process design of hybrid routing protocols in considering real-time mobile nodes.

## III. MOBILE COMMERCE SYSTEM

The adhoc mobile commerce takes place between multiple numbers of nodes which are in proximity of each other without relying on the services of any infrastructure [29] which is very different from infrastructure mobile commerce application. Such nodes can be termed as peer node which can cooperate and participate in communication process by using their normal local resources along with their neighbor's independent on any support provided by a network service provider in order to achieve the transaction or such related task. So, following are the inherent properties of mobile commerce in adhoc network:

- *Independent of Service Provider*: As the adhoc wireless network will not have a network service infrastructure and are self-organized, a dedicated service provider cannot be entrusted for allocation of maintenance task for enabling security parameters or payment scheme reliability for m-commerce applications.

- *Restricted Scope of Communication*: With various restriction in communication range especially in IEEE 802.11 [30] [31] [32] poses a challenges in adhoc networks where the network topology is normally dynamic rendering less trust on any third party service by communicating peer node to support security and/or payment in real-time application among the peers.

- *Inadequate online time*: Due to finite energy cycle and the dynamic topology of mobile devices as well as intermittent network disconnections [33], there is a restricted time during which these mobile devices can be online, which actually limits them from participating in lengthy and complex operation processes related to transactions [34]. This fact represents that complex secure operation in m-commerce need to be completed in a fairly short period and should only comprise a few simple stages if they are to have a good chance of success in terms of security.

- *Impulsive choice in Adhoc configurations*: As the adhoc wireless network has self-organizing attribute [33] which allows client that are equipped with mobile devices to instinctively participate in m-commerce transactions when the requirement arises while they are on the mobile mode.

- *Cost Effectiveness*: There is no extra complicated device [35] in mobile commerce application in order to perform security operations in m-commerce over an adhoc wireless network as peer nodes which will formulate the network.

- *Privacy*: The mobile commerce application enabled in adhoc wireless network is very much appropriate for maintaining or safeguarding the privacy protocol for commercial transactions where the clients (traders) will not look for disclosing the commercial transaction information to some external entities [36]. There is no third entity which needs to be involved in order to realize the network communication.

## IV. APPLICATIONS

There is numerous distinctive variety of m-commerce secure transactional processes that can be conducted out over adhoc wireless networks:

- *Digital Resources Quality*: Exchange of digital contents such as music files, eBooks, videos, etc. For example, two peers' nodes who meet by chance at an terminal may be in agreement to exchange a digital contents for a specific reason.

- *Auction in M-Commerce*: An auction process [37] can be designed anywhere as soon as a group of at least three peer nodes with mobile devices and shared software are in mutual agreement to participate in trading process. Such type of activity is agreeable to short term involvement by peer nodes and a rapid turnover in its membership providing enough are frequently present to create a significant volume of bidders. Multicasting among participant can distribute bids and information about what is information related to an offer [38].

- *Amusement-Products*: With the increase of interactive gaming products and such types of products among diminutive groups of community is another breed of application appropriate to adhoc networking. Certain function running on mobile devices considers the products, manages its communications and handle the turnover in participants [39].

- *Machine Enabled Transactions*: Various online transactions which use mobile devices that are already equipped with electronic-cash in order to make compensation at a point of sales (POS) and so on via IEEE 802.11 technologies [40].

- *Private Transactions*: In the dynamic environment of mobile adhoc network, it might happen that two or more peers can meet on the go and then agree to exchange their private information definitely with making aware of such transactions to any third party agents [41] in the wireless network

- *Collective Trading*: In mobile adhoc network, there is a huge feasibility that a group or communities of peer nodes who are in transmission range of each other and facilitated with mobile devices with each other and can spontaneously formulate a group for collective trading.

- *Electronic IOUs*: 'I Owe U' or its abbreviation 'IOU' is an established means to acknowledge a small debt usually among friends or family members [42]. This form of acknowledgement can be passed electronically via an ad hoc wireless network among trading parties who are in close proximity with each other. It can be signed to verify its authenticity and the identities of all handling parties.

## V. PROBLEM DESCRIPTION

The serious issues to be highlighted are that conducting m-commerce secure operations over mobile adhoc wireless networks introduces added challenges and concern. Along with this, the adhoc wireless networks have particular problems which needed to be considered in research works in future. The major issued found are illustrated as below:

- *Transaction Management*: It is very difficult to execute secure effective transaction methods and moreover updates in mobile adhoc networks, which is due to its sole distinctiveness e.g. lack of infrastructure, having a dynamic network topology and using resource constrained devices. Majority of the traditional research work has utilized infrastructure based m-commerce which depends on a client/server model where information is fundamentally located placed on servers within the wired network and peer nodes act as clients accessing the services provided by the servers [43] along with an issue of service unavailability due to network disconnections [43]. Also, the in-depth idea of a transaction can be difficult to enforce as network intermittent disconnections will affect a particular service in a secure m-commerce operations succession to fail and accordingly the secure connectivity would be considered unfinished and will be subjected to abort [44].

- *Delivery of Service*: Due to the unique characteristics and complexities of an adhoc wireless network, existing service discovery and delivery protocols [45] do not seem to suit the needs of an adhoc wireless network, making them unsuitable for m-commerce oriented scenarios. Service advertisements and deliveries may need to be disseminated by a mix of a store and forward strategy as well as local multicasting to cope with intermittent online connectivity [45].

- *Trust-System*: One of the important factor of online communication in terms of security will be Trust, which assists in participating entities to ensure the secure transaction by extenuating improbability and risks involved in the transactions, such as ambiguity about trading groups or entities' pattern in fulfilling the transaction agreements [46]. On the other hand, as mobile adhoc network cannot rely on a network service provider to facilitate security services such as certification authority (CA) which can assists to design trust system among peer nodes in the existing network. It can also be observed that peer nodes have to rely on their peers in the network to provide trust verification in order to evaluate other nodes' fidelity. Yet, the nature of an adhoc wireless network makes trust scheme founding in this network intricate to accomplish.

The problems encountered in this field is multiplied more as to date, the majority of adhoc routing protocol research has been done using simulation only. One of the most motivating reasons to use simulation is the difficulty of creating a real implementation. In a simulator, the code is contained within a single logical component, which is clearly defined and accessible [47]. On the other hand, creating an implementation requires use of a system with many components, including many that have little or no documentation. The implementation developer must understand not only the routing protocol, but all the system components and their complex interactions. Results obtained through simulations cannot be ascertained for standardization of efficient of routing protocols in challenging environment of MANET in real time scenario, which will definitely pose an obstacle for real-time research and development. Further, since adhoc routing protocols are significantly different from traditional routing protocols, a new set of features must be introduced to support the routing protocol. In this journal, we describe the event triggers required for performing AODV operation, the design possibilities and the decisions for our Ad hoc On-demand Distance Vector (AODV) routing protocol implementation, AODV-DSDV. This journal is meant to aid researchers in developing their own on-demand mobile ad hoc routing protocols and assist users in determining the implementation design that best fits their real-time needs.

## VI. PROPOSED SYSTEM

The main approach of the proposed design is to create a design for process oriented mobile commerce framework in real test bed. In order to do so, the prominent focus is laid to the design of the effective routing protocols as the experimental test bed is real time where the user are mobile and their mobility is very much variable, which makes the framework more challenging to ensure the effective security towards the architecture.

One of the great advantages of AODV is its integrated multicast routing [49]. Routing protocols are categorized into two types depending on how and when routes are discovered, but both discover the shortest path to the destination. Proactive routing protocols are table-driven protocols which always maintain current up-to-date routing information by transmitting control messages periodically between the hosts which update their routing tables. When there are alterations in the structure then the updates are propagated using the network. The proactive routing protocols deploy link-state routing algorithms which frequently flood the link information about its neighbors. Other routing protocols are on-demand routing protocols, in other words reactive one which create routes when they are needed by the source host and these routes are maintained while they are required. Such protocols use distance-vector routing algorithms; they have vectors containing information about the cost and the path to the destination. When nodes exchange vectors of information, each host modify own routing information when required. Till date, the majority of adhoc routing protocol research has been done using simulation only as seen in the previous section of related work. One of the most motivating reasons to use simulation is the difficulty of creating a real implementation. It is required that the implementation developer must comprehend not only the routing protocol, but all the system components and their complex interactions. Furthermore, since adhoc routing protocols are significantly different from traditional routing protocols, a new set of features must be introduced to support the routing protocol, which is the main focus of this research journal.

Some components where an adhoc wireless network like mobile ad hoc network (MANET) or Bluetooth is considered as a medium to carry out the transactions[49]. As the mobile commerce transactions will normally involve only mobile devices which are peers and have no assurance of infrastructure support from a network service provider, there are only two crucial entities involved in proposed adhoc m-commerce.

- Client: This entity is mainly mobile and utilizes adhoc wireless networks in order to get the digital contents [50], goods or services offered by the vendor or to business contents, products or services for others.

- Vendor: This entity provides the digital contents, products or services directly to clients via adhoc wireless networks for capital or who makes business out of the contents, products or services for others. In such condition, the vendor may also be stationary for example, a vending machine. Nonetheless, there are diversified feasible vital entity relationships in the adhoc m-commerce value chain.

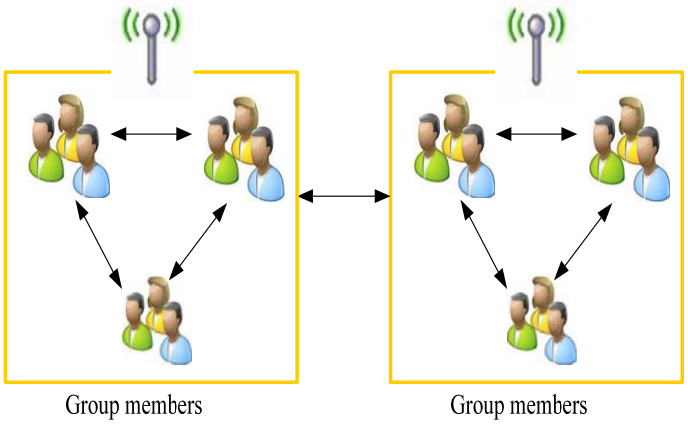



Fig.2: Trading between two consortiums

.

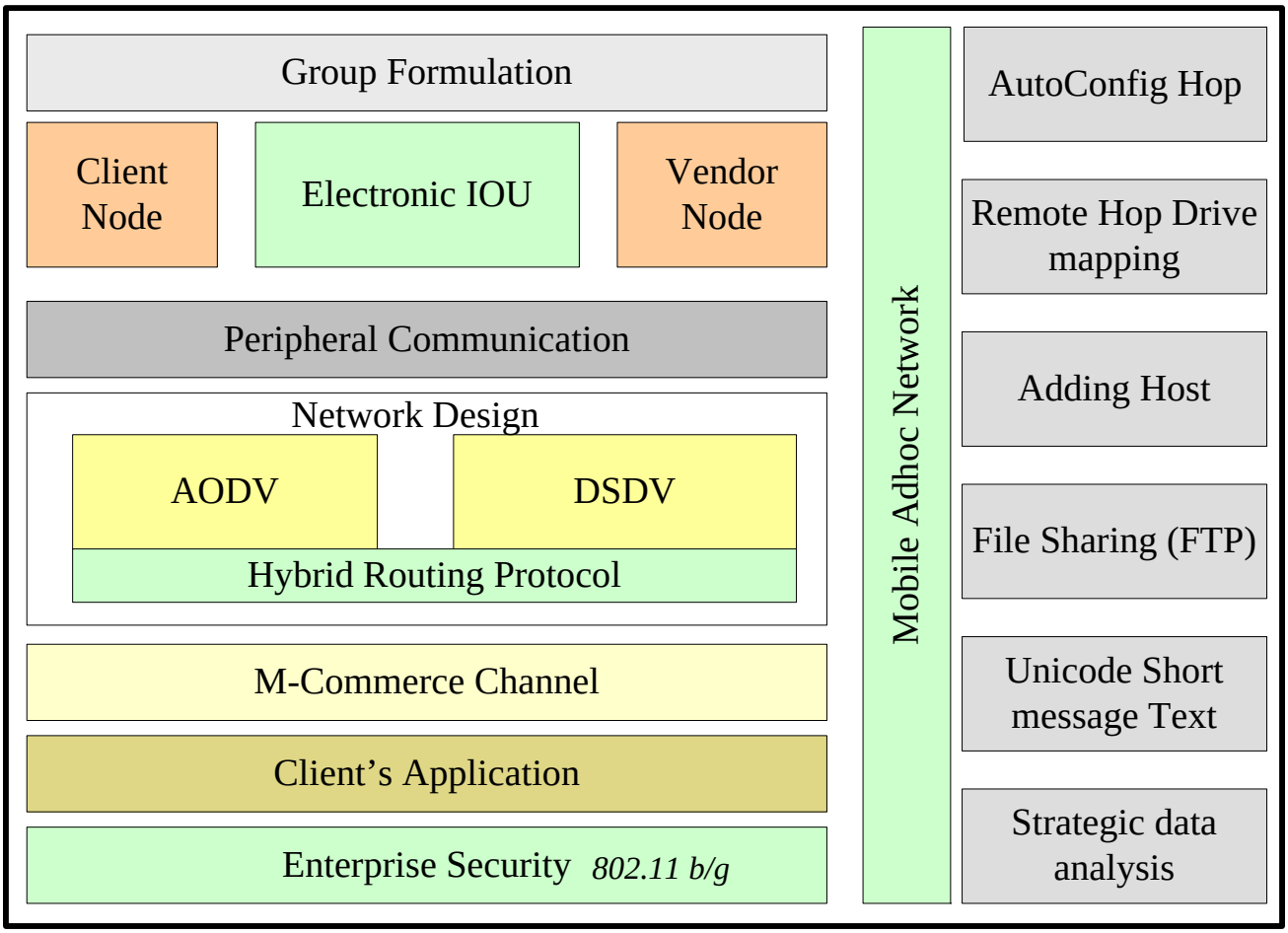



Fig 3. Process Oriented framework of the proposed m-commerce system

Fig 3. Represents the process oriented framework of the proposed mobile commerce system. The design is a consideration of the creating a secure mobile transaction in overlay network of our proposed hybrid protocol using AODV as well as DSDV.

The network will consists of peer nodes such as mobile devices or smart phones can also be utilized in order to purchase digital contents from a vendor, or initiate imbursement at specified Point of Sales (POS). E-cash credits can be preloaded and used via a Bluetooth or Wi-Fi technology to communicate with the vending machine, POS or parking toll [40]. But in most of the complex environment, where there can be multiple peer nodes may be involved in the secure m-commerce operations like auctions

Figure 2 represents a comparatively simple transaction involving two peer nodes. For example, two entities who are commuting in a medium be in agreement to trade their e-digital contents while they are within radio range of each other.

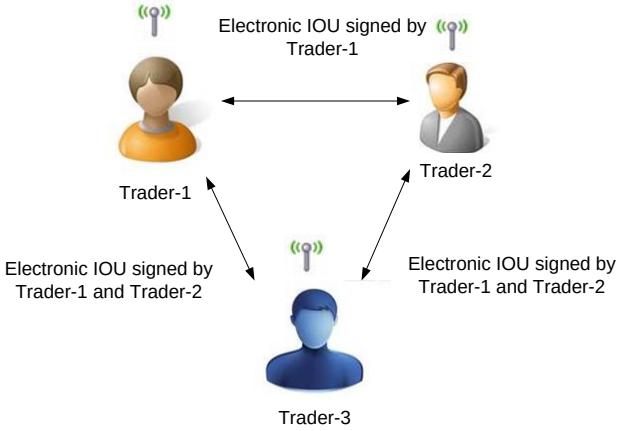



Fig.4. A delegated trading scenario.

The network highlights dual environments which involves the configuration of an adhoc trading consortium among mobile users who are in the surrounding area of each other and agree to group together for a explicit function, for example to make a collective purchase or to involve in a group trading. Figure 4 represents a hand over trading environment where an

electronic I Owe U (IOU) [42] is used to recognize debt between two parties trading via an ad hoc wireless network.

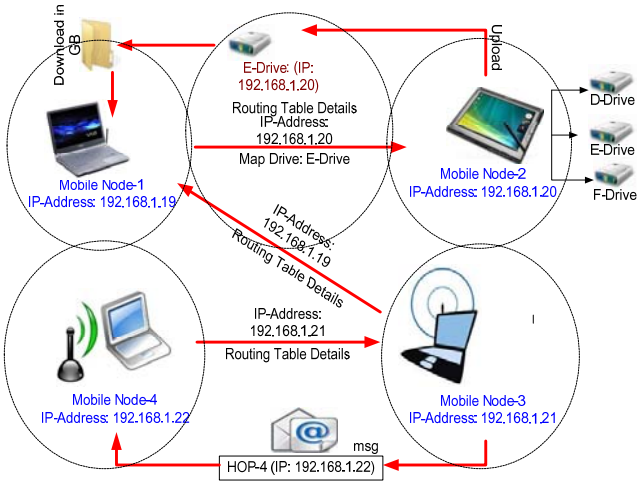



Fig.5. Overview of the real time set up plan

As an example, figure 4 below highlights a scenario in a local neighborhood where Trader 1, who has a digital product, wants to trade it for an another digital product. Trader 2, who is assumed to be within Trader 1's communication range and actually owns an digital product, agrees to trade with Trader 1 but does not want a have the digital product owned by Trader-1. So, Trader 1 issues an electronic IOU securely signed by himself to Trader 2 as an acknowledgement of his debt to Trader 2. Trader 2 can afterward exercise that electronic IOU to trade for another digital contents such that she wants with Trader 3, who wants a another digital content. Trader 3 will then use the electronic IOU signed by Trader 1 and Trader 2 to settle with Trader 1 for his product.

The goal of this research work is a real-time implementation of an adhoc routing protocol, using the 802.11 standard wireless protocols. Our implementation enables communication between several wireless stations or hops, on a dynamic network without using any infrastructure, i.e. using peer-to-peer mode, rather than depending on access points. Two distant units can communicate even when there is no direct connection between them.

The project is planned to implement two potential algorithms:

- Direct Sequence Distance Vector (DSDV) algorithm, which is a pro-active table driven algorithm. The routing in each station is executed according to local routing table. The tables are continually maintained and updated.

- Ad-hoc on demand Distance Vector (AODV) algorithm, which is a reactive algorithm, which operates only when there is demand from upper layer to send data.

The experiment is designed in Java platform, which has inherent support for wireless network operations. Thus, it is platform independent, and can run with various OS and IEEE 802.11 wireless cards. The scheme of real time set up is as highlighted in Fig 5.

## VII. Experimental Setup

In order to perform this experiment, 5 Motorola Xoom Tablet PC, 5 IBM laptop and 10 Acer laptop with Intel Atom processor with 1.80 GHz, Ram Size of 2039 MB, and 32 bit OS. The wireless configuration will be 802.11a/b/g Wi-Fi inbuilt. The prototype design we are researching on is a windows application which runs on the wireless adhoc network. The project is planned to implement two potential algorithms: Direct Sequence Distance Vector (DSDV) algorithm, which is a pro-active table driven algorithm. The routing in each station is executed according to local routing table. The tables are continually maintained and updated. Ad-hoc on demand Distance Vector (AODV) algorithm, which is a reactive algorithm, which operates only when there is demand from upper layer to send data. We chose to write the project in Java, which has inherent support for network operations. Thus, it is platform independent, and can run with various OS and wireless cards. Other functionalities of the project work includes auto configuration of hops for Middleware, Remote Hop Drive Mapping and host adding, File sharing with FTP with statistical transfer ratio and time detection, Transfer of Unicode Short Message Text to the destination hop, Strategic data analysis of Hop sequence, metric sequence number, lifetime, type, Detecting configuration with respect to header information, size of data packets, buffer size when reading file. The nodes move from a random starting point to a destination with a speed ranging from 0-5 m/sec.

## VIII. Performance Analysis

The proposed process oriented architecture of the mobile commerce system depends completely on the efficient design of our routing protocol. The protocol needs to be checked for sustaining the highly dynamic mobility of the experiment in terms of packet delivery ratio. In order to check the performance of the real-time experimental test-bed, the criteria of evaluation will be to check the packet delivery ratio of large files normally in terms of GB along with bandwidth estimation. The application is also designed to estimate the bandwidth for checking the QoS. Fig 6. Comparison of packet delivery ratio of routing schemes. The proposed hybrid scheme is compared with traditional AODV, DSDV, DSR routing protocols by considering packet delivery ratio as the parameter for comparison along with number of mobile nodes (10). Packet delivery ratio was estimated by estimating

number of packet received by packets sent. The final estimates of the packet delivery ratio was confirmed using commercial network protocol analyzer (WireShark).

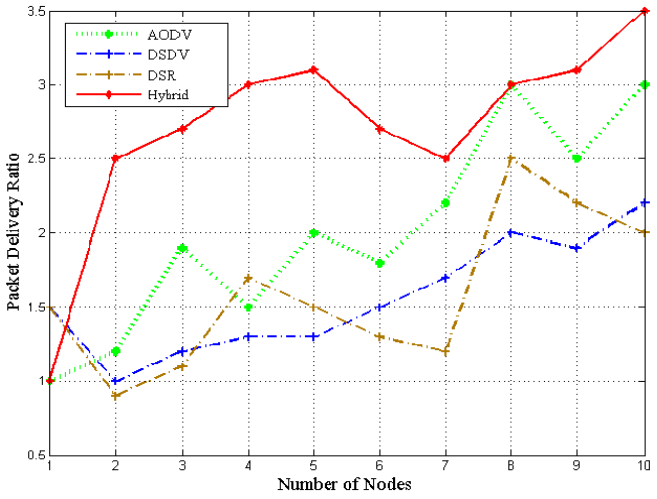



Fig 6. Comparison of packet delivery ratio of routing schemes

The proposed hybrid protocol as well as AODV has better results with increasing number of mobile nodes increasing packet delivery ratio. The packet delivery ratio of AODV is found to be increased as soon as quantity of mobile nodes as well as the pause time of the mobile nodes movement are incremented. DSDV has the better synchronous output till 8th node where the observation concludes as a very gradual rise, but performance has little degradation when the new mobile node (9th) joins the network, but however, it manages to communicate smoothly after 9th mobile node. The performance of the DSR protocol is found to be not so smooth with increment of mobile nodes. Majority of the experiment were performed in a closed room free from any EM radiation or any types of interference. In order to check the seamless connectivity, the 6th-10th mobile nodes were asked to conduct movement away from the closed room or in parking lawn but within sensing range of 5 meters. File of bigger size (GB) as well as peer activity application (transmission of Unicode short text character) has been attempted to transmit to check the efficiency of hybrid protocol, which was found to be work effectively.

## IX. Security Requirements

In order to establish an appreciably tenable and trusted atmosphere for an operation to take place as well as to facilitate self-reliance to trading entities to participate in secure m-commerce operations, the following security services can be considered as an important functional requirement.

i. *Authentication*: Authentication is the first step which facilitates both trading entities participated in m-commerce transactions to substantiate the identity of each other before any transaction is conducted among the groups. This service ensures that any illegal third party or external agents is not masquerading as a legitimate party.

ii. *Privacy*: Privacy guarantees that secure transaction information sent across the network is incomprehensible by illegal third parties such as malicious eavesdroppers or peers acting only as communication relays, or DDoS attacks.

iii. *Reliability*: Reliability ensures that a message is being transferred is not illegitimately altered or destroyed during the transmission without this being detectable at the receiver side of an m-commerce system.

iv. *Non-repudiation*: This is another property of which assures that if an entity transmits a message, it will not be able to move away with denying after sending the message. Usually in m-commerce transactions, neither sender nor receiver should realistically be able to renounce offers or bargains struck between them. The sender should not be able convincingly to deny having sent the transaction message and the receiver should be able to verify that the transaction message can only have been sent by the specified sender and thus able to prove that a business has taken place between them. Along with this, as m-commerce transactions will include the threat of misbehavior among the trading entities, which they need support in measuring the intensity of reliability of other trading entities. Hence, attestation is another important security service for an adhoc m-commerce.

v. *Attestation*: Attestation enables an adhoc m-commerce peers to vouch for the identity, trading history or transaction reputation of other peer nodes. It assists alleviate threat in transacting with formerly unknown entities.

## X. Conclusion

The proposed research journal discusses about the real time test-bed by designing process oriented architecture of a new hybrid protocol combining AODV and DSDV. The experimental results shows satisfactory results when the communication is attempted for larger size of file from one to another hop in mobile adhoc network. Developing a working implementation of an adhoc routing protocol is non-trivial and more difficult than developing a simulation. In real time implementation, various factors like battery lifetime, interference present in the room where the experiment is conducted, presence of EM radiations, mobility of nodes, performance of mobile device etc affects the results, when estimating the communication among the mobile nodes. The

simulation results conducted in the previous research work cannot be followed as a standard guidelines for commercial usage as the user might have dynamic scenario of implementation in real-time, which might not have been considered in simulation test bed. Once the programmer designs the application, then at the time of evaluation in practical scenario, various dynamic scenarios are often found missing at the time of coding process like fluctuation of battery, optimal performance of the laptops, which cannot be predicted about their performance, thereby posing a great challenge in implementing hybrid protocols in real time test bed. For these reasons it takes significantly more effort to create an ad hoc routing protocol implementation than a simulation. The wireless business as seen on mobile commerce networks seems most appropriate to fully online resource exchange and also to online launch trading in local groups where entities could easily meet to transfer digital contents and payment as agreed. But as seen in this research journal, in order to support such types of mobile commerce is more challenging as compared to wireless commerce within provider networks. This research journal highlights entities which will act as guidelines for serving the understanding for better quality efficient, secure, and reliable m-commerce system. There can be further work in terms of experimenting the communication in with respect to vehicular adhoc network, by estimating the average delay, packet drop, and packet delivery ratio, which will be our focus in future work. In summary, adhoc networks have the potential to become a serious part of tomorrows 4G communications networks. They can open up new business opportunities for network operators and service providers.

References

[1] Anup K. Ghosh, Tara M. Swaminatha, Software security and privacy risks in mobile e-commerce, Magazine Communications of the ACM CACM Homepage archive Volume 44 Issue 2, Feb. 2001

[2] Dr. C. Rajabhushanam and Dr. A. Kathirvel, “Survey of Wireless MANET Application in Battlefield Operations”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.1, January 2011

[3] Dr. Jyotsna Sengupta and Er. Gurpreet Singh Grewal, “Performance evaluation of IEEE 802.11 MAC layer in supporting delay sensitive services”, International Journal of Wireless & Mobile Network (IJWMN), Vol 2, No. 1, Feb 2010.

[4] http://www.fixedwirelessnetwork.com/ [Access on 7th July, 2011]

[5] T. Ravi Nayak, Dr. K Ashok Babu, “Implementation of adaptive Zone routing protocol for Wireless networks”, International Journal of Engineering Science and Technology, Vol. 2 (12), 2010,

[6] http://www.ietf.org/rfc/rfc2501.txt [Access on 7th July, 2011]

[7] Maiya, S.V. Fuja, T.E., “One-Hop vs. Two-Hop Routing in Simple Networks with Fading: An Outage Probability Analysis Addressing Spectral Efficiency”, Wireless Communications and Networking Conference, 2008.IEEE

[8] Yang Lei FengJun Shang Zhaohua Long Yunsen Ren, “An Energy Efficient Multiple-Hop Routing Protocol for Wireless Sensor Networks”, Intelligent Networks and Intelligent Systems, 2008.

[9] http://www.livinginternet.com/i/iw_route_igp_ospf.htm [Access on 7th July, 2011]

[10] Brian James Wolf, “Cross-layer scheduling protocols for mobile ad hoc networks using adaptive direct-sequence spread-spectrum modulation”, Doctoral Thesis, 2009

[11] Xiaoyan Hong; Kaixin Xu; Gerla, M., “Scalable routing protocols for mobile adhoc networks”, Network, IEEE Issue Date: Jul/Aug 2002

[12] Tao Lin, “Mobile Ad-hoc Network Routing Protocols: Methodologies and Applications”, Doctoral Thesis, 2004

[13] Aleksandr Huhtonen, “Comparing AODV and OLSR Routing Protocols”, IEEE 2004

[14] Charles E. Perkins and Pravin Bhagwat, “Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers,” In Proceedings of the SIGCOMM '94 Conference on Communications Architectures, Protocols and Applications, pages 234--244, August 1994

[15] Thomas Clausen, Philippe Jacquet, Anis Laouiti, Pascale Minet, Paul Muhlethaler, Amir Qayyum, and Laurent Viennot, “Optimized Link State Routing Protocol”, Internet-Draft, draft-ietf-manet-olsr-06.txt, September 2001

[16] S. Murthy and J. J. Garcia-Lunes-Aceves, "An Efficient Routing Protocol for Wireless Networks," ACM Balzer Mobile Networks and Application Journal, vol. 1, no. 2, pp. 183--197, November 1996

[17] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networking," in Mobile Computing, T. Imielinski and H. Korth, Eds. Norwell, MA: Kluwer, 1996.

[18] Charles Perkins and Elizabeth Royer, “Ad hoc on demand distance vector routing” In Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pages 90--100, Feb 1999.

[19] Zygmunt J. Haas, “A new protocol for the reconfigurable wireless networks,” School of Electrical Engineering, Cornell University, Ithaca, NY, 14853 http: // www.ee.cornell.edu/~haas/wnl.html

[20] Nicklas Beijar, "Zone Routing Protocol (ZRP)," Networking Laboratory, Helsinki University of Technology PO Box 3000, FIN-02015 HUT, Finland.

[21] Yang Cheng-yun, Zhang Ming-qing, Tang Jun, “Design of Secure Efficient Routing Scheme Based on AODV”, Computer Engineering, Doi： CNKI:SUN:JSJC.0.2010-01-056, 2010

[22] Dr. Upena Dalal D Rakesh Kumar Jha Suresh V. Limkar. "Article:A Performance Comparison of Routing Protocols(DSR and TORA) for Security Issue In MANET(Mobile Ad Hoc Networks)". IJCA Special Issue on MANETs (2):78–83, 2010. Published by Foundation of Computer Science

[23] Shukla, A.K., Tyagi, N., "A new route maintenance in dynamic source routing protocol", Wireless Pervasive Computing, 2006 1st International Symposium, Issue Date: 16-18 Jan. 2006

[24] Javad Akbari Torkestania, Mohammad Reza Meybodi, "Mobility-based multicast routing algorithm for wireless mobile Ad-hoc networks: A learning automata approach", Computer Communications Volume 33, Issue 6, 15 April 2010, Pages 721-735

[25] Rafael Martínez-Peláez, Francisco Rico-Novella, Cristina Satizábal and Jhon J. Padilla, Performance Analysis of Mobile Payment Protocols over the Bluetooth Wireless Network, IEEE 2008

[26] Suresh Chari, Parviz Kermani, Sean Smith, and Leandros Tassiulas, Security Issues in M{Commerce: A Usage{Based Taxonomy, Springer-Verlag Berlin Heidelberg 2001

[27] Alia Fourati and Khaldoun Al Agha, Secure and fair auctions over ad hoc networks, Int. J. Electronic Business, Vol. 5, No. 3, 2007

[28] Osman, Husna; Taylor, Hamish, managing group membership in adhoc e-commerce trading system, New Technologies of Distributed Systems (NOTERE), 2010 10th Annual International Conference on 2010

[29] Dipanjan Chakraborty, Harry Chen,, Service discovery in the future for mobile commerce, Magazine Crossroads Crossroads Homepage archive Volume 7 Issue 2, December 2000
ACM New York, NY, USA

[30] Barnes, S.J., Under the Skin: Short-range Embedded Wireless Technology. International Journal of Information Management, 2002. 22(2002): p. 165-179.

[31] Tiwari, R., S. Buse, and C. Herstatt. The Mobile Commerce Technologies: Generations, Standards and Protocols. 2006 [cited 20/06/08]; Hamburg University of Technology, Institute of Technology and Innovation Management Available from: http://www1.uni-hamburg.de/mcommerce/articles/Working_Paper_40.pdf. p. 1-21.

[32] Levy, R.; Carlos, P.S.; Teittinen, A.; Haynes, L.S.; Graff, C.J.; Mobile agents routing-a survivable ad-hoc routing protocol, Military Communications Conference, 2005. MILCOM 2005. IEEE

[33] De Pellegrini, F.; Miorandi, D.; Carreras, I.; Chlamtac, I.; A Graph-Based Model for Disconnected Ad Hoc Networks, INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE

[34] Mike Spreitzer, Marvin Theimer, Providing location information in a ubiquitous computing environment (panel session), Proceeding SOSP '93 Proceedings of the fourteenth ACM symposium on Operating systems principles ACM New York, NY, USA ©1993

[35] Nansi Shi, Mobile commerce applications, Idea Group Inc (IGI), 2004 - 344 pages

[36] Ranjan B. Kini and Somboon Thanarithiporn, Mobile commerce and electronic commerce in Thailand: a value space analysis, International Journal of Mobile Communications Issue: Volume 2, Number 1 / 2004 Pages: 22 – 37

[37] Rosemary Stockdale, Craig Standing, (2002) "A framework for the selection of electronic marketplaces: a content analysis approach", Internet Research, Vol. 12 Iss: 3, pp.221 – 234

[38] Michael Christoffel, Jens Nimis, Bethina Schmitt, Peter C. Lockemann, An Infrastructure for an Electronic Market of Scientific Literature, 2000

[39] Jo Groebel, Eli M. Noam, Valerie Feldmann, Mobile media: content and services for wireless communications, Routledge, 2006 - 255 pages

[40] Woodings, R.W.; Joos, D.D.; Clifton, T.; Knutson, C.D.; Rapid heterogeneous ad hoc connection establishment: accelerating Bluetooth inquiry using IrDA, Wireless Communications and Networking Conference, 2002. WCNC2002. 2002 IEEE

[41] Minch, R.P., Privacy issues in location-aware mobile devices, System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on 2004

[42] Mohd. Sahandri Gani B. Hamzah1, Mohd. Reza Ghorbani1, Saifuddin Kumar B. Abdullah2, The impact of electronic communication technology on written language, Nov. 2009, Volume 6, No.11 (Serial No.60)

[43] Perich, F., et al., Neighborhood-Consistent Transaction Management for Pervasive Computing Environment, in 14th International Conference on Database and Expert Systems Applications (DEXA 2003). 2003, UMBC Ebiquity Research Group. p. 1-10

[44] Jianguo Ding; Balasingham, I.; Bouvry, P.; Management challenges for emerging networks and services, Ultra Modern Telecommunications & Workshops, 2009. ICUMT '09. International Conference on 2009

[45] Helal, S.; Desai, N.; Verma, V.; Choonhwa Lee;, Konark - a service discovery and delivery protocol for ad-hoc networks, Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE

[46] Patil, V. and S.R. K, Trust Management for E-Transactions, in Sadhana. 2005, Indian Academy of Science. p. 141-158

[47] Ian Chakeres Elizabeth M. Belding-Royer. "AODV Implementation Design and Performance Evaluation." International Journal of Wireless and Mobile Computing (IJWMC) Issue 2/3, 2005

[48] Rainer Baumann, "Adhoc On-Demand Distance vector Routing Protocol", Accessed from http://www.baumann.info/public/aodv.pdf

[49] Amitabh Mishra, Security and quality of service in ad hoc wireless networks, Cambridge University Press, 2008 - Technology & Engineering - 180 pages
[50] Hartung, F.; Ramme, F.; , Digital rights management and watermarking of multimedia content for m-commerce applications, Communications Magazine, IEEE 2000

# Genetic Algorithm and Confusion Matrix for Document Clustering

**A. K. Santra, C. Josephine Christy,**
**[1] Dean, CARE School of Computer Applications , Trichy – 620 009, India.**

**[2] Research Scholar, Bharathiar University, Coimbatore – 638401,India.**

**Abstract**

Text mining is one of the most important tools in Information Retrieval. Text clustering is the process of classifying documents into predefined categories according to their content. Existing supervised learning algorithms to automatically classify text requires sufficient documentation to learn exactly. In this paper, Niching memetic algorithm and Genetic algorithm (GA) is presented in which feature selection an integral part of the global clustering search procedure that attempts to overcome the problem of finding optimal solutions at the local less promising in both clustering and feature selection. The concept of confusion matrix is then used for derivative works, and finally, hybrid GA is included for the final classification. Experimental results show benefits by using the proposed method which evaluates F-measure, purity and results better performance in terms of False positive, False negative, True positive and True negative.

**Keywords: Text mining, GA, Confusion matrix, F-measure**

## 1. Introduction

In Text data mining, Text classification has become one of the most important techniques. The task is to automatically classify documents into predefined classes based on their content. Many algorithms have been developed to deal with document clustering. With the existing algorithms, a number of newly established processes are involving in the automation of Document clustering. It has been observed that for the purpose of Document clustering the concept of association rule is very well known. Association rule mining finds interesting association or correlation relationships among a large set of data items. The discovery of these relationships among huge amounts of transaction records can help in many decision making process. On the other hand, the confusion matrices use the maximum a posteriori estimation for learning a classifier. It assumes that the occurrence of each word in a document is conditionally independent of all other words in that document given its class.

The confusion matrix is more commonly named contingency table in which the matrix could be arbitrarily large. The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified accurately. Improved Genetic algorithm starts with an initial population which is created consisting of randomly generated rules. Each rule can be represented by a string of bits. Based on the notion of survival of the fittest, a new population is formed to consist of the fittest rules in the current population, as well as offspring of these rules. Typically, the fitness of a rule is assessed by its classification accuracy on a set of training examples.

This paper presents an improved genetic algorithm which is used to evaluate the weights of the metrics such as F-measure, purity and accuracy. We apply improved genetic algorithm to find out and identify the potential informative features combinations for classification and then use the F-Measure to determine the fitness in genetic algorithm. The improved GA is general purpose search algorithm which provides rules inspired by natural genetic populations to evaluate solutions to problems. In our method, not as usual, an individual is joined together of the real-coded metrics' weight, and it's more natural to indicate the optimization problem in the continuous domain.

### 2. Literature Review

A. K. Santra, C. Josephine Christy and B. Nagarajan [1] have proposed that cluster based niche memetic and genetic algorithm have been designed & implemented by optimizing feature selection of text in the document repository. The contribution of genetic algorithm works with an evaluation of fitness function. Accuracy can be calculated through the document clustering. S. Areibi and Z. Yang [2] have proposed several local search operations to effectively design an MA for simultaneous clustering and feature selection. which incorporate local searches with traditional GAs,

have been proposed and applied successfully to solve a wide variety of optimization problems. These studies show that pure GAs are not well suited to fine tuning structures in complex search spaces and that hybridization with other techniques can greatly improve their efficiency. S. Wu *et al.*[3] have proposed about data clustering is a common technique for statistical data analysis and has been used in a variety of engineering and scientific disciplines such as biology (genome data). Y. Zhao and G. Karypis [5] have proposed the purity of a cluster represents the fraction of the cluster corresponding to the largest class of documents assigned to that cluster; thus, the purity of the cluster.

One way of approaching this challenge is to use stochastic optimization schemes, prominent among which is an approach based on genetic algorithms (GAs). The GA is biologically inspired and embodies many mechanisms mimicking natural evolution. It has a great deal of potential in scientific and engineering optimization or search problems. Recently, hybrid methods [8], which incorporate local searches with traditional GAs, have been proposed and applied successfully to solve a wide variety of optimization problems. These studies show that pure Gas [16] are not well suited to finetuning structures in complex search spaces and that hybridization with other techniques can greatly improve their efficiency. GAs that have been hybridized with local searches are also known as memetic algorithms (MAs) [7].

Traditional GAs and MAs are generally suitable for locating the optimal solution of an optimization problem with a small number of local optima. Complex problems such as clustering, however, often involve a significant number of locally optimal solutions. In such cases, traditional GAs and MAs cannot maintain controlled competitions among the individual solutions and can cause the population to converge prematurely [3]. To improve the situation, various methods [7], (usually called niche methods) have been proposed. The research reported shows that one of the key elements in finding the optimal solution to a difficult problem with a GA approach is to preserve the population diversity during the search, since this permits the GA to investigate many peaks in parallel and helps in preventing it from being trapped in local optima. GAs are naturally applicable to problems with exponential search spaces and have consequently been a significant source of interest for clustering [6, 10]. For example, in [4] proposed the use of traditional GAs for partitioned clustering. These methods can be very expensive and susceptible to becoming trapped in locally optimal solutions for clustering large data sets.

In [8] introduced hybrid GAs by incorporating clustering-specified local searches into traditional GAs. In contrast to the methods proposed in [11] and [12], clustering based on hybrid GAs can be more efficient, but these techniques can still, however, suffer from premature convergence. Furthermore, all of the above methods may exhibit limited performance, since they perform clustering on all features without selection. GAs have also been proposed for feature selection [7]. However, they are usually developed in the supervised learning context, where class labels of the data are available, and the main purpose is to reduce the number of features used in classification while maintaining acceptable classification accuracies. The second (and related) theme is feature selection for clustering, and feature selection research has a long history, as reported in the literature.

Feature selection in the context of supervised learning, adopts methods that are usually divided into two classes filters and wrappers based on whether or not feature selection is implemented independently of the learning algorithm. To maintain the filter/wrapper distinction used in supervised feature selection, we also classify feature selection methods for clustering into these two categories based on whether or not the process is carried out independently of the clustering algorithm [13, 14, 15]. The filters in clustering basically preselect the features and then apply a clustering algorithm to the selected feature subset. The principle is that any feature carrying little or no additional information beyond that subsumed by the remaining features is redundant and should be eliminated.

## 3. Document Clustering

While document clustering can be valuable for categorizing documents into meaningful groups, the usefulness of categorization cannot be fully appreciated without labeling those clusters with the relevant keywords or key phrases that describe the various topics associated with them. A highly accurate key phrase extraction algorithm, called Core Phrase is proposed for this particular purpose.

Core Phrase works by building a complete list of phrases shared by at least two documents in a cluster. Phrases are assigned scores according to a set of features calculated from the matching process. The candidate phrases are then ranked in descending order and the top L phrases are output as a label for the cluster. While this algorithm on its own is useful for labeling document clusters, it is used to produce cluster summaries for the collaborative clustering algorithm.

Document clustering is used to organize a large document collection into distinct groups of similar documents. It discerns general themes hidden within the corpus. Applications of document clustering go beyond organizing document collections into knowledge maps. This can facilitate subsequent knowledge retrievals and accesses. Document clustering, for example, has been applied to improve the efficiency of text categorization

and discover event episodes in temporally ordered documents. In addition, instead of presenting search results as one long list, some prior studies and emerging search engines employ a document clustering approach to automatically organize search results into meaningful categories and thereby support cluster-based browsing.

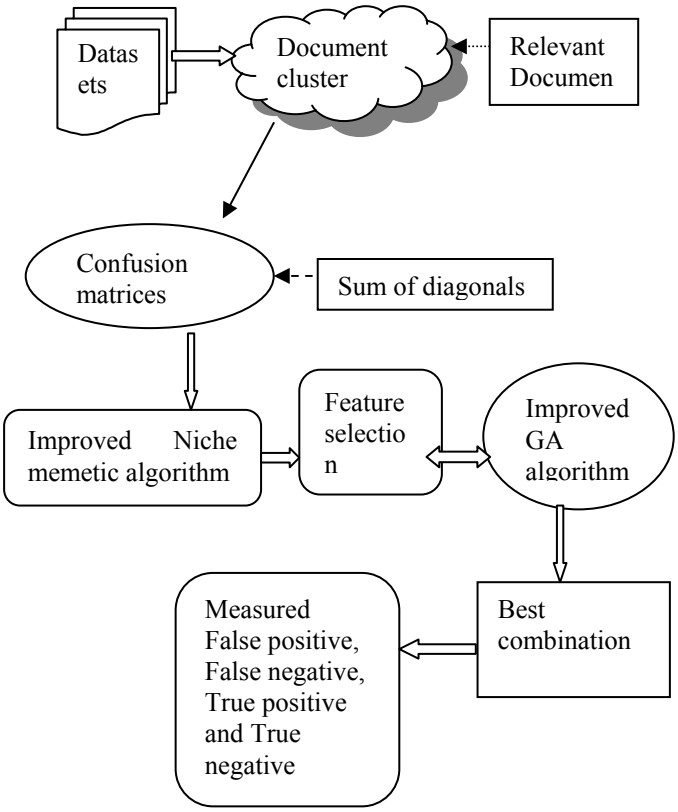



**Fig 1 : Document Clustering using confusion matrices on Improved GA algorithm**

## 4. Feature Selection

Feature selection is important for clustering efficiency and effectiveness because it not only condenses the size of the extracted feature set but also reduces any potential biases embedded in the original (i.e., non-trimmed) feature set . Previous research commonly has employed feature selection metrics such as TF (term frequency), TF×IDF (term frequency × inverse document frequency), and their hybrids.

Unlike the non-LSI-based document clustering approach, which typically involves a feature selection phase, the LSI-based approach to clustering monolingual documents employs LSI to reduce the dimensions and thereby improve both clustering effectiveness and efficiency. Its process generally commences with feature extraction, followed by document representation.

### 4.1 Confusion Matrix

A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

| | | Predicted | |
|---|---|---|---|
| | | Negative | Positive |
| Actual | Negative | a | b |
| | Positive | c | d |

The entries in the confusion matrix have the following meaning in the context of our study:

- a is the number of correct predictions that an instance is negative,
- b is the number of incorrect predictions that an instance is positive,
- c is the number of incorrect of predictions that an instance negative, and
- d is the number of correct predictions that an instance is positive.

Several standard terms have been defined for the 2 class matrix:

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$AC = \frac{a + d}{a + b + c + d} \quad \text{----------> (1)}$$

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c + d} \quad \text{-----> (2)}$$

- The false positive rate (FP) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

$$FP = \frac{b}{a + b} \quad \text{------> (3)}$$

- The true negative rate (TN) is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$TN = \frac{a}{a + b} \quad \text{-------> (4)}$$

- The false negative rate (FN) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

$$FN = \frac{c}{c + d} \quad \text{-------> (5)}$$

- Finally, precision (P) is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$P = \frac{d}{b + d} \quad \text{-------> (6)}$$

The accuracy determined using equation 1 may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases. Suppose there are 1000 cases, 995 of which are negative cases and 5 of which are positive cases. If the system classifies them all as negative, the accuracy would be 99.5%, even though the classifier missed all positive cases. Other performance measures account for this by including TP in a product: for example, geometric mean (g-mean), as defined in equations 7 and 8, and F-Measure (Lewis and Gale, 1994), as defined in equation 9.

$$\text{g-mean}_1 = \sqrt{TP * P} \quad \text{--------> (7)}$$

$$\text{g-mean}_2 = \sqrt{TP * P} \quad \text{--------> (8)}$$

$$F = \frac{(\beta 2 + 1) * P * TP}{B2 * P + TP} \quad \text{--------> (9)}$$

In equation 9, b has a value from 0 to infinity and is used to control the weight assigned to TP and P. Any classifier evaluated using equations 7, 8 or 9 will have a measure value of 0, if all positive cases are classified incorrectly.

### 4.2. Niching Memetic Algorithm

One of the key elements in overcoming less promising locally optimal solutions of a difficult optimization problem with a GA approach is to preserve the population diversity during the search. In this section, we introduce a modification of the niching method and integrate it into our GA to preserve the population diversity during the simultaneous search for clustering and feature selection.

The niching method presented was designed for clustering where no feature selection is required and the number of clusters is known beforehand. In this method, a niching selection with a restricted competition replacement was developed to encourage mating among similar solutions while allowing for some competitions among dissimilar solutions. The flow of the algorithm is given as follows:

**Step 1**: Initialize the population_size p
**Step 2:** For each p in initial population, p = local search (p)
**Step 3**: Calculate unified criterion for each of the offspring. If the fitness of the offspring is better than its paired solution, then the latter is replaced.
**Step 4:** Provide the feature subset and cluster centers of the solution from the terminal population with the best fitness.

### 4.3. GA Algorithm

A genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. The flow of the algorithm is given as follows:

**Input:** document set DS, number of generations n
**Output:** best classifier over DS

**Step 1:** Evaluate the sets of candidate positive and negative terms
**Step 2:** Create the population oldPop and initialize each chromosome
**Step 3:** Evaluate the fitness of each chromosome in oldPop
**Step 4:** Copy in NewPop the best r chromosomes of oldPop

**Step 5**: While size(newPop) < size(oldPop)
- select parent1 and parent2 in oldPop
- generate kid1, kid2 through crossover(parent1, parent2)
- apply mutation, i.e., kid1 = mut(kid1) and kid2 = mut(kid2)
- apply the repair operator ρ to both kid1 and kid2
- add kid1 and kid2 to newPop

**step 6:** oldPop = newPop
- Select the best chromosome K in oldPop;
- Eliminate redundancies from K;

**step 7:** classifier associated with K.

### 5. Performance Evalution

The performance of improved GA on Documents is evaluated in this section. Let us suppose that we have obtained a clustering solution with feature selection. Since the quality of clusters depends on the particular application, there is no standard criterion for evaluating

clustering solutions. We compute classification errors, since we know the "true" clusters of the synthetic data and the class labels of the real data. This is done by first running the algorithm to be tested on each data set. Next, each cluster of the clustering results is assigned to a class based on examining the class labels of the data objects in that cluster and choosing the majority class. After that, the classification errors are computed by counting the number of misclassified data objects. For the identification of correct clusters, initially we report the number of clusters found. We stress that the class labels are not used during the generation of the clustering results, and they are intended only to provide independent verification of the clusters.

The feature recall and precision are reported on synthetic data, since the relevant features are known a priori. Recall and precision are concepts from text retrieval. Feature recall is the number of relevant features in the selected subset divided by the total number of relevant features. Feature precision is the number of relevant features in the selected subset divided by the total number of features selected. These indices give us an indication of the quality of the features selected. High values of feature recall and precision are desired. Note that, with respect to the real data, we report only the number of feature selected, since the relevant features are unknown.

## 6. Experimental Result and Discussion

The proposed method was tested with a file of 100 historical documents. The datasets were taken as related topic of Data mining, Image processing and Networking. For each dataset, 30% of the documents are randomly selected as test documents, and the rest are used to create training sets as follows: γ percent of the documents from the positive class is first selected as the positive set P. The rest of the positive documents and negative documents are used as unlabeled set U. We range γ percent from 10%- 50% to create a wide range of scenarios.

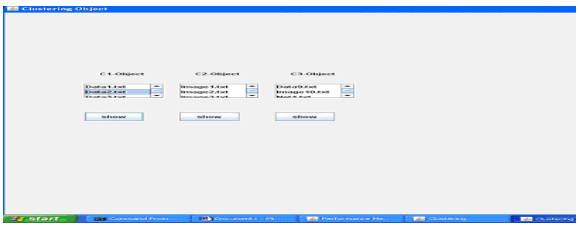


Fig 2 Clustered DataSet

Preliminarily, documents were subjected to the following pre-processing steps: (1) First, we removed all words occurring in a list of common stopwords, as well as punctuation marks and numbers; (2) then, we extracted all n-grams, defined as sequences of maximum three words consecutively occurring within a document (after stopword removal)5; (3) at this point we have randomly split the set of seen data into a training set (70%), on which to run the GA, and a validation set (30%), on which tuning the model parameters. We performed the split in such a way that each category was proportionally represented in both sets (stratified holdout). Based on the term frequency and inverse document frequency, the term weight will be calculated.

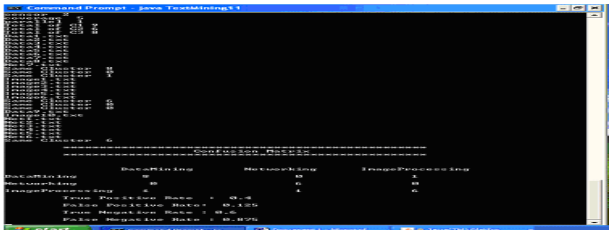


Fig 3: Confusion matrix on text documents

The performance results are measured in terms of F-measure, purity and false positive rate according to the Number of documents and Cluster object.

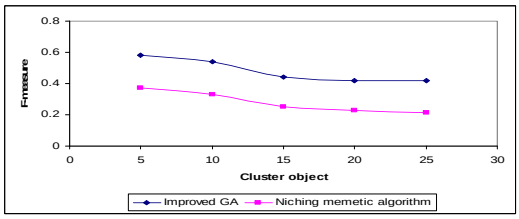



Fig 4: Cluster object vs F-measure

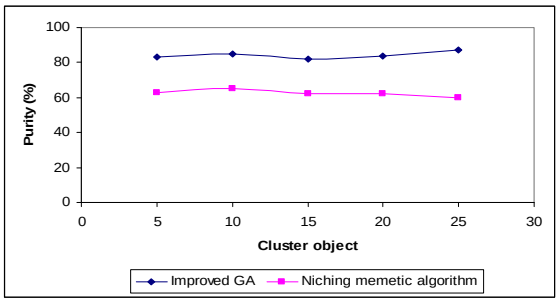



Fig 5: Cluster vs purity

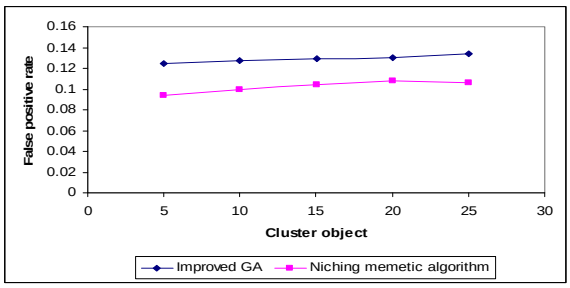



Fig 6: Cluster Object vs False positive rate

Figure (4), (5) and (6) shows the result of F-measure, purity and false positive rate with respect to the cluster object. The proposed algorithm improved GA gives the better result compared with existing method, Niching memetic algorithm.

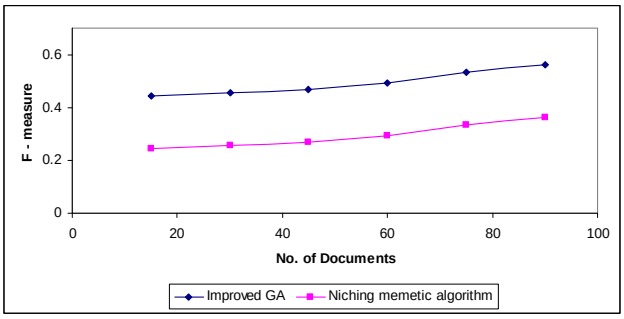



Fig 7: No. of Documents vs F-measure

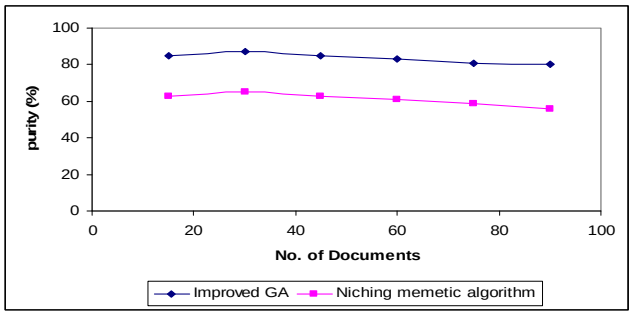



Fig 8: No. of Documents vs Purity

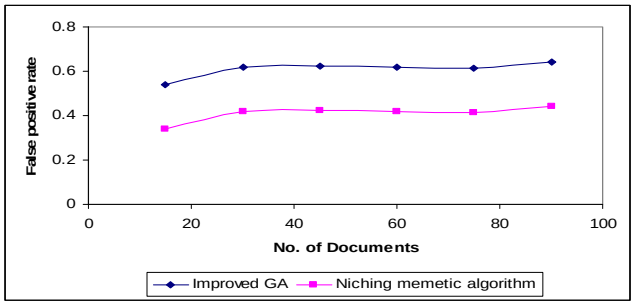



Fig 9: No. of documents vs false positive rate

Figure (7), (8) and (9) depicts the performance result of F-measure, purity and false positive rate according to the Number of documents. It is observed that improved GA performs the well. By comparing niching memetic algorithm with improved GA, proposed improved GA can efficiently recover solutions with low classification errors

## 6. CONCLUSION

The improved Niche memetic algorithm and improved genetic algorithm have been designed and implemented by using confusion matrices. Our proposed method is applied to real data sets with an abundance of irrelevant or redundant features. Improved GA relies on confusion matrices and uses the F-measure as the fitness function. In this case, identifying a relevant subset that adequately captures the underlying structure in the data can be particularly useful. Additionally, as a general optimization framework, the proposed algorithm can be applied for text mining. In such a case, an unbiased clustering criterion in some sense is produced by computing the mutual information between clusters, thus enabling a better verification of the properties of the proposed optimization scheme. We conclude by remarking that we consider the experimental results can further be improved through a fine-tuning of the GA parameters.

## References

[1] A.K. Santra, C. Josephine Christy and B.Nagarajan, "Cluster Based Hybrid Niche Memetic and Genetic Algorithm for Text Document Categorization", IJCSI, vol.8, Issue 5, no. 2,pp. 450-456, Sep 2011.

[2] S. Areibi and Z. Yang, "Effective Memetic Algorithms for VLSI Design Automation = Genetic Algorithms + Local Search + MultiLevel Clustering," Evolutionary Computation, vol. 12, no. 3, pp. 327- 353, 2004.

[3] S. Wu, A.W.C. Liew, H. Yan, and M. Yang, "Cluster Analysis of Gene Expression Database on Self-Splitting and Merging Competitive Learning," IEEE Trans. Information Technology in Biomedicine, vol. 8, no. 1, 2004.

[4] H.K. Tsai, J.M. Yang, Y.F. Tsai, and C.Y. Kao, "An Evolutionary Approach for Gene Expression Patterns," IEEE Trans. Information Technology in Biomedicine, vol. 8, no. 2, pp. 69-78, 2004.

[5] Y. Zhao and G. Karypis, "Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering," Machine Learning, vol. 55, no. 3, pp. 311-331, 2004.

[6] J. Kogan, C. Nicholas, and V. Volkovich, "Text Mining with Information-Theoretic Clustering," IEEE Computational Science and Eng., pp. 52-59, 2003.

[7] W. Sheng, A. Tucker, and X. Liu, "Clustering with Niching Genetic K-Means Algorithm," Proc. Genetic and Evolutionary Computation Conf. (GECCO '04), pp. 162-173, 2004.

[8] K. Deep and K. N. Das. Quadratic approximation based Hybrid Genetic Algorithm for Function Optimization. AMC, Elsevier, Vol. 203: 86-98, 2008.

[9] C. Wei, C.S. Yang, H.W. Hsiao, T.H. Cheng, Combining preference- and content-based approaches for improving document clustering effectiveness, Information Processing & Management 42 (2) (2006) 350–372.

[10] Renchu Guan, Xiaohu Shi, Maurizio Marchese, Chen Yang, and Yanchun Liang, "Text Clustering with Seeds Affinity Propagation" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011

[11] Y.J. Li, C. Luo, and S.M. Chung, "Text Clustering with Feature Selection by Using Statistical Data," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 5, pp. 641-652, May 2008.

[12] B.J. Frey and D. Dueck, "Non-Metric Affinity Propagation for Un- Supervised Image Categorization," Proc. 11th IEEE Int'l Conf. Computer Vision (ICCV '07), pp. 1-8, Oct. 2007.

[13] L.P. Jing, M.K. Ng, and J.Z. Huang, "An Entropy Weighting KMeans Algorithm for Subspace Clustering of High-Dimensional Sparse Data," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 8, pp. 1026-1041, Aug. 2007.

[14] Z.H. Zhou and M. Li, "Distributional Features for Text Categorization," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 3, pp. 428-442, Mar. 2009.

[15] F. Pan, X. Zhang, and W. Wang, "Crd: Fast Co-Clustering on Large Data Sets Utilizing Sampling-Based Matrix Decomposition," Proc. ACM SIGMOD, 2008.

[16] Jung-Yi Jiang, Ren-Jia Liou, and Shie-Jue Lee, "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 3, MARCH 2011

**A. K. Santra** received the P. G. degree and Doctorate degree from I.I.T., Kharagpur in the year 1975 and 1981 respectively. He has got 20 years of Teaching Experience and 19 years of Industrial (Research) Experience. His area of interest includes Artificial Intelligence, Neural Networks, Process Modeling, Optimization and Control. He has got to his credit (i) 35 Technical Research Papers which are published in National / International Journals and Seminars of repute, (ii) 20 Research Projects have been completed in varied application areas, (iii) 2 Copy Rights for Software Development have been obtained in the area of Artificial Neural Networks (ANN) and (iv) he is the contributor of the book entitled **"Mathematics and its Applications in Industry and Business"**, Narosa Publishing House, **New Delhi.** He is the recognized Supervisor for guiding Ph. D. / M. S. (By Research) Scholars of Anna University-Chennai, Anna University-Coimbatore, Bharathiyar University, Coimbatore and Mother Teresa University, Kodaikanal. Currently he is guiding 12 Ph. D. Research Scholars in the Department. He is a Life member of CSI and a Life member of ISTE.

**C. Josephine Christy** received her M.Sc., M.Phil., M.B.A., from Bharathiar University, Coimbatore. Currently she is working as Asst.Professor in Bannari Amman Institute of Technology, Sathyamangalam. Her area of interest includes Text Mining, Web Mining. She presented a paper in International Journal, 2 papers international conferences and 6 papers in national Conferences. She is a Life member of Computer Society of India and a Life member of Indian Society for Technical Education.

# Implication of Secure Micropayment System Using Process Oriented Structural Design by Hash chaining in Mobile Network

Chitra Kiran N.
*Research Scholar*
Dept of Electronics & Communication Engg.
UVCE
Bangalore, India

Dr. G. Narendra Kumar
*Prof.* Dept. of Electronics & Communication Engg.
UVCE
Bangalore, India

***Abstract*— The proposed system presents a novel approach of designing a highly secured and robust process oriented architecture for micropayment system in wireless adhoc network. Deployment of any confidential transaction over dynamic nature of wireless adhoc network will strike a high amount of security challenges which is very difficult to identify which poses a great difficulty in designing and effective countermeasures. The current work designs the security process using hash chain and Simple Public Key Infrastructure to be implemented on newly designed digital agreement of broker along with paving new secure routing for secure m-transaction as an efficient alternative for digital coin. The system stimulates the intermediate nodes to cooperate for facilitating secure and reliable transaction from source to destination nodes. The system consists of high end encryption using hash function is also independent of any Trusted Third Party when the network topology frequency changes, thereby it is flexible, lightweight, and reliable for secure micropayment systems. The analysis result shows the system is highly robust and secure ensuring anonymity, privacy, non-repudiation offline payment system over wireless adhoc network.**

***Keywords-Micropayment, M-Commerce, hash-function, offline transaction***

## I. INTRODUCTION

Wireless adhoc network has become one of the prime topics of research in the very recent years where majority of the research work is concentrated on restricted user-groups, where various nodes cooperate to communicate [1]. But security and energy consumption is always a never ending issue in wireless adhoc network. Although wireless adhoc network can be effectively used in wireless payment system cost effectively, but unfortunately, such technology comes with many security flaws. One of the prominent classes of payment found to be used in m-commerce recently is micropayment system [2] which is based managing small payment values. Mobile payment is defined as the process of exchanging financial values between two parties using a mobile device to pay for products or services [3]. With this new payment option, customers can pay for products and services anywhere and anytime with the comfort offered by their mobile devices. It is designed to operate with wireless technologies such as Bluetooth, Infrared or 802.11x[4]. The electronic payment system over the wireless mobile adhoc network is one of the considerable topics of research currently. Such type of network is characterized by dynamic topology, unwanted energy consumption, and obvious link breakage. Therefore creating a dedicated and secure payment system of ubiquitous type will become a very challenging task for any researchers. From the decentralized and infrastructureless types of the network, various threats might evolved due to dynamic topology caused by random mobility of the device as well as restricted resources on trusted handheld devices. In anonymous micropayment schemes, there is no connection between the payer and the payment means. In this case, the payment means should be secured by a third party vendor which is normally any financial institutions. The financial institution should ensure the reliability and the legitimacy of each coin in the network which also means that every user who wants to verify a coin should check with the financial institution. The second type of payment is in connection to the payer, where each payment mean or token should include the characteristics of the first payer. Therefore, before accepting any payment mean, a node should substantiate the first payer and verify that he owns requires the involvement of a trusted third party. Not only this, but the payee can directly redeem the payment means or use the similar token for another payment, if the micropayment mechanism allows asking for a delegation authorization. Commercially various e-payment system are in use which works on cellular network [5], but the success rate is very low due to high security threats. Majority of existing transaction systems are online and are directly dependent on a fixed cellular network with increased cost for service. Such system currently in practice has no assurity of reliability and exposed a privacy infringes implicating threats to payment systems. While electronic commerce (e-

commerce) continues to have a profound impact on the global business environment, technologies and applications have begun to focus more on mobile computing and the wireless Web. With this trend comes a new set of issues and problems specifically related to wireless e-commerce. Ultimately, researchers and developers must determine what tasks users really want to perform anytime from anywhere and decide how to ensure that information and functionality to support those tasks are readily available and easily accessible [6] The communications infrastructure necessary for the wireless Internet environment is quite complex. Wireless devices are likely to remain at a disadvantage over their wired counterparts in terms of bandwidth. Limited bandwidth is a significant problem that requires organizations to rethink how users interact through a wireless device with an information system. An important issue is how to create efficient applications that can realistically work with current technology [6]. Accordingly, micropayment schemes still requires the proper designing of efficient security protocols, which could become problematical according to the quantity of the payers and the environment of the payment means and payment chains. Further, this system does not describe any robust mechanisms allowing to conclude distributed payment or pay distributed applications.

Abundant researches for e-payment system have been already proposed [7][8][9]. The researches on payment system over mobile network have been discussed in [10]. Such system has extensive deployment of expensive cryptographic protocol operations. Micropayment systems has contributed to iterative payments from a single vendor where majority of the security policies has used one-way hash functions [11] in order to generate a chain of hash values. Hash functions such as MD5/SHA are more computationally proficient in comparison to other symmetric key algorithms such as AES or asymmetric key algorithms such as RSA and allow for fast generation and verification of payment tokens [12]. But maximum of the researches comes with a security loopholes and high costing. Use of advance cryptographic protocols in such cases will only increase the memory and network overhead for high requirement of maintenance of key management. So traditional cryptography cannot be deployed in securing the communication between one to another node in wireless adhoc network. The problem of reliability of communication becomes much worst when there is a frequent changes in the network topology. This paper provides an overview of some of the relevant technologies, applications, and issues in the relatively new field of wireless e-commerce.

This research paper will provide solution for accomplishing secure and flexible e-payment system over wireless mobile adhoc network. The proposed system does not consider any online transaction like traditional system but it is designed for offline e-payment system over wireless mobile adhoc network using Simple Public Key Infrastructure (SPKI) [13]. The system integrates almost all the banking needs very securely and cost-effectively with well adaption to direct deposits, e-cheque, amount transfer etc eradicating the threats of exposing private e-payment information to illegal third party. This phenomenon will make man-in-middle attack or any unauthorized user very difficult to explore the location of effective attack as there is no central entity in the transaction path over wireless adhoc network. The proposed system considers distributed authorization controls for various modules, where one module delegates to the other about the permission.

In Section II, we will discuss about the previous research work in this area followed by Section III about electronic payment scheme in Wireless Adhoc Network. Section IV highlights processing of Micropayment system followed by proposed system Discussion in section V. In depth discussion of research methodology is done in section-VI followed by description of architecture in section-VII. Section VIII highlights the performance analysis of the conducted experiment followed by security requirement analyzation and conclusion in section-X

## II. RELATED WORK

Zhi-Yuan Hu [14] has designed an innovative and practical authentication system, Anonymous Micropayments Authentication (AMA), is designed for micropayments in mobile data network. But his work has a relative drawback for common problems of authentication mechanism based on symmetric key cryptography.

Xiaoling Dai [15] has researched on micropayment protocols in offline with multiple vendors.

Min-Shiang [16] has introduce several micro-payment schemes based on one-way hash chain and review some literatures on supporting multiple payment. The author has also proposed a new micropayment scheme, which achieve the following three goals: micro-payment multiple transactions, service providers, and anonymity.

Samad [17] has proposed a trust model from user point of view and combined it with MR2 micropayment scheme and called the new scheme TMR2. This trust model is supported by micropayment provider and assures the users that they will not be charged for in case the product is not satisfactory or it is corrupt.

Sung-Ming e.t. al [18] has studied various probabilistic micropayment Scheme shows that the scheme by Rivest may reduce the administrative cost of the bank, however it brings extensive computational overhead to the merchant.

Lih-Chyau Wuu [19] has proposed a secure and efficient off-line micro payment scheme which uses coin chain technique to make coin that the verification of coin can be done quickly by hash computation. This scheme also ensures that the

coins could only be used by their owner, and protects the privacy of the consumer.

Vivek Katiyar e.t. al. [20] has discussed about role of Elliptical Curve Cryptography and presents a survey on the current use of ECC in the pervasive computing environment.

Husna Osman and Hamish Taylor [21] has discussed three key design considerations in implementing a fully distributed reputation system for ad hoc m-commerce trading systems, namely relevant reputation information, its storage and reliability.

Fouzia Mousumi and Subrun Jamil [22] has described cost effective push pull services officering SMS based mobile banking concept has been illustrated for 24 hours banking convenience which helps customers stay on top of any recent changes made in their current or deposit account or loan through SMS.

Arogundade e.t. al. [23] propose an open network system which can adapt to users changing needs as well as allowing effective and secured transaction via any customers' bank account.

Partha e.t. al [24] proposes a novel approach by utilizing cancelable biometric features for securely storing the fingerprint template by generating Secured Feature Matrix and keys for cryptographic techniques applied for data Encryption or Decryption.

Mohammad Al-Fayoumi [25] discuss an important e-payment protocol namely pay-word scheme and examine its advantages and limitations, which encourages the authors to improve the scheme that keeps all characteristics intact without compromise of the security robustness

Kaylash Chaudhary e.t. al [26] have carried out an assessment of micro-payment against a non-micro-payment credit systems for file sharing applications.

Charles K. Ayo and Wilfred Isioma Ukpere [27] propose a unified (single) smart card-based ATM card with biometric-based cash dispenser for all banking transactions

Wang [28] proposes a novel payment system with smart mobile devices, wherein customers are not limited to purchase e-cash with the fixed face-value

Currently, researchers focus on the e-payment system such that electronic cash [29-34], electronic check [35, 36], electronic traveler's check [37][38] and so on.

Moreover, many researchers proposed the e-cash payment protocol [29-34], using plenty of computational resources such that exponential operation. It causes the big burden for the system. Chang and Lai [33] proposed a flexible date-attachment scheme on e-cash and Juang [35] proposed the D-cash. Curan [39] introduces some possible additional security measures which could be implemented to strengthen the overall security architecture of Bluetooth enabled devices for m-commerce applications against man-in-the middle attack and denial-of-service attacks.

Wang e.t. al [40] proposes a novel payment system with smart mobile devices, wherein customers are not limited to purchase e-cash with the fixed face-value. The amount of every transaction is deducted directly from the customer's account, eliminating the inconvenience of fixed face-value of the e-cash, and reducing online computation cost of a bank. Using a technique of trapdoor hash function to mitigate the computational cost, our system can be used with the mobile devices effectively.

Natarajan [41] introduced a system and method of extensible authentication protocols (EAPs) based on ECC and SKE with a permutation technique evolved. The permutation in our EAPs is a process of cubing a random number w.r.to a prime. These EAPs are compatible with 3G and 4G networks and no certificates exchanged during the communication.

Panjwani [42] has analyzed two token-based authentication schemes, designed for authenticating users in banking systems implemented over mobile networks. The first scheme is currently deployed in India by a mobile banking service provider named Eko with a reach of over 50,000 customers. The second scheme was proposed recently (in joint effort with Eko) to fix weaknesses in the first one, and is now being considered for deployment. Both systems rely on PINs and printed codebooks (which are unique per user) for authentication. Chaix [43] explores the economic models associated to different mobile-payment systems.

Obviously it can be seen that majority of the work is carried on wired network with much less consideration of wireless network. The issues related to dynamic topologies of wireless adhoc network is not discussed in detailed in any of the researches described above. Although there are some effective research being done in the area of payment system, but there is a huge research gap in this area with respect to wireless mobile adhoc network.

## III. SCHEMA OF E-PAYMENT

According to the definition of mobile payment in mobile payment forum, mobile payment is the financial transactions for some services or good between the trading parties through mobile terminals [44]. Businessmen or service providers can transfer the regulated electronic money from their own mobile-phone-bound account to other accounts through mobile phone, with the assistance of mobile payment environment providers [44]. In comparison with online payment, the mobile payment consists of one more responsibility, namely mobile payment service provider, which is a significant position in the whole payment performance since the trading could only be completed with the vigorous cooperation of mobile communication operators mainly due to the insecurity of mobile payment and the immaturity of this field [45]. Mobile payment makes it available to conduct trading anytime and anywhere, which is the biggest advantage of this mode of payment [46]. But this system also comes with various lethal security threats posing a greatest challenge in designing a secure payment system in wireless adhoc network. Majority of the payment system currently in use consider online communication with the network and is much infrastructure

dependent, which is very different scene compared to wireless mobile adhoc network. The use of digital coin is also in abundant. But it has been seen that digital coin usage generates security issues as well as privacy issues. The pre-requisites of deployment of effective and secure payment systems in wireless mobile adhoc network are as follows:

- The security of the session can be ensured working on offline mode as direct access to central server is impossible.
- The system must ensure anonymity for the user thereby protecting the real identity of user involved in the system.
- Avoid dual payment in one transaction.
- Avoid forged or illegal resources..
- Ensuring non-repudiation for the user involved, the vendors, and the bank
- Increased efficacy must be guaranteed for optimal usage of memory and resources involved..
- The system should not use much advanced hardware for deployment in order to reduce the complexity involved in maintaining security
- The system must be scalable.

## IV. MICROPAYMENT SYSTEM

A micropayment is a financial transaction involving a very small amount of money and usually one that occurs online [47]. One problem that has prevented their emergence is a need to keep costs for individual transactions low which is impractical when transacting such small sums even if the transaction fee is just a few cents [47]. Micropayments have to be appropriate for the transaction of non-tangible merchandise over the Internet which inflicts necessities on speed and cost of processing of the payments: delivery occurs nearly immediately on the Internet, and often in arbitrarily small pieces. On the other hand, the bottleneck in sales of tangible merchandise, management and distribution, sets a lower bound particularly for costs to remain economical. So, the evaluation criteria of micropayment systems should include [48]:

- *Ease of use*: The application must be easy to use for the clients. There is no authorization login and PIN number to be fed all the time. The customer only needs to click and to buy a page in the web page with a micropayment system in a few seconds.
- *Security*: The aim of security in the payment procedures is to prevent any group from cheating the system. For customers and external adversaries the forms of cheating security, which are detailed to payment design, are extra expenditure of coins and creation of false coins forgery during payment.
- *Anonymity*: The customer anonymity should be protected. An elementary property of physical cash is that the association between customers and their purchases is untraceable. This means that the payment systems do not allow payments to be traced without compromising the system's security. This may encourage some potential customers to start using the payment system.
- *Divisibility:* The protocol supports multiple denominations and a range of payment values.
- *Performance:* The protocol provides high-volume payment support.
- *Robustness:* The protocol is tolerant of network bottlenecks and broker/authorizer down-time.

Table 1. Comparison of E-commerce payment methods

| **Property** | **CyberCash [48]** | **MPay [48]** | **PayWord [48]** | **NetPay [48]** |
|---|---|---|---|---|
| Ease of Use | Low | High | Medium | High |
| Security | High | Medium | Low | Medium+ |
| Anonymity | Low | Low | Low | Medium+ |
| Divisibility | Very High | Very High | High | High |
| Performance | Very Low | High | Medium | Very High |
| Robustness | Low | High | High | High |

There is a growing need for an effective, efficient micro-payment technology for high-volume, low-value E-commerce products and services. Current macro-payment approaches do not scale to such a domain. Most existing micro-payment technologies proposed or prototyped to date suffer from problems with security, lack of anonymity and performance

## V. PROPOSED SYSTEM

The proposed system is based on the secure and reliable transaction being carried out in an offline connection in wireless adhoc network. The proposed system is standardized with respect to communication system where it facilitates ease in deployment for clients. The proposed model will amalgamate into the hierarchical transaction system which facilitates the clients to conduct transaction both in online as well as in offline connection with reliable security measures. The proposed architecture (See Figure 2) is designed for security features using simple public key infrastructure which integrates elasticity to the use of e-cheque in offline mode using digital coins. The proposed system highlights a secure micropayment system by which the system allocates a payment to all those nodes which permits relaying of the packets thereby providing service. Such types of the nodes implicate the payment agreement to pay. The payment agreement can be governed along with the uniqueness of each node in the mobile network. This can also be verified by the Trusted Third Party (TTP). Unfortunately wireless adhoc network will not support these long-lived service (payment) agreements among the nodes due to the dynamic topology of the wireless adhoc

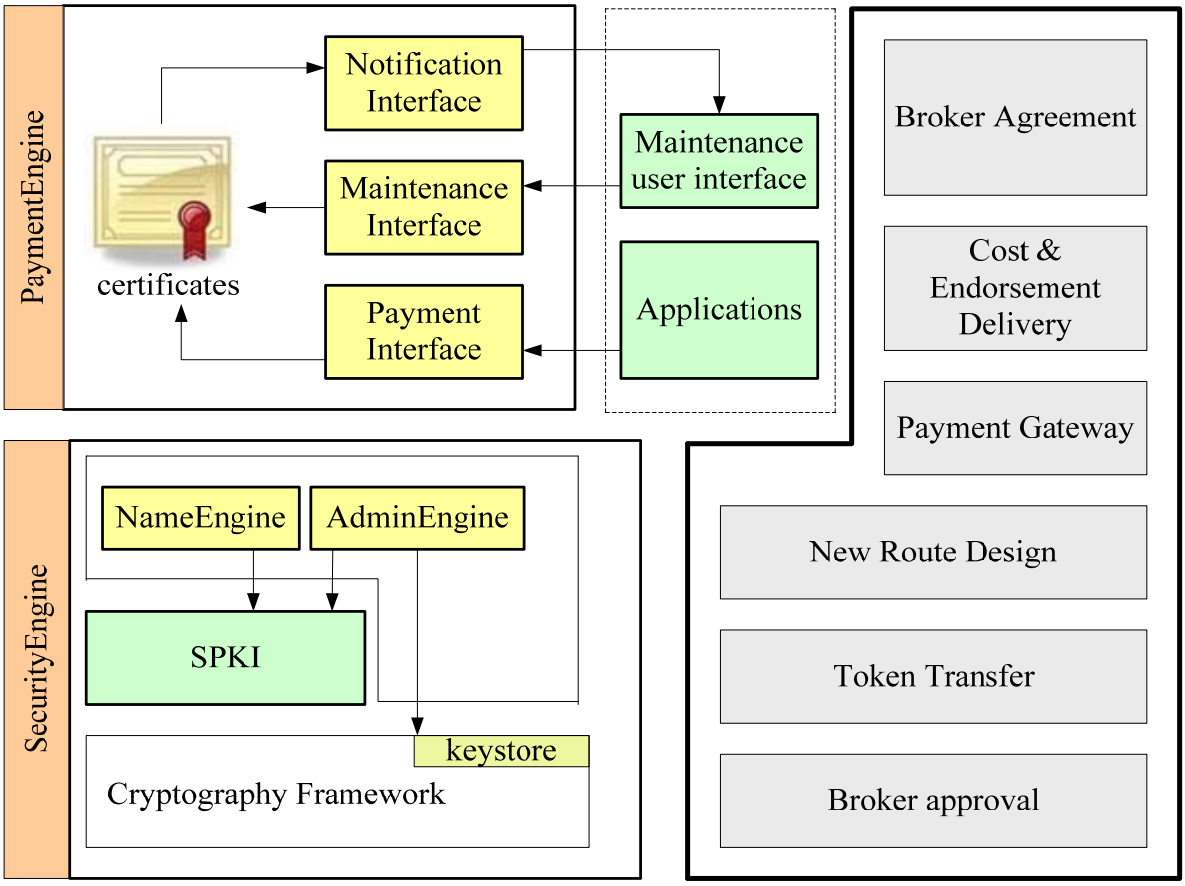



Fig.1. Architecture of the proposed structural design

network, where it is very difficult to predict the position of the nodes in next sequence of time. Therefore, there is a need of extensible as well as secure policy which allows the user to make payment to all nodes in the network without any dependency on TTP or any financial institutions to issue a new payment agreement.

The main aim of the research work is to design a secure protocol which stimulates the nodes for packet forwarding in wireless adhoc network. The objective of the proposed research journal are:

- *Verification*: The system should allow both online as well as offline validation of the payment tokens independent from any need of intermediate relay nodes.
- *Route Flexibility*: The scheme should permit selection of an most favorable route towards its destination and initiate payment to all nodes in its network. In case of route diversion, the system is independent from TTP to create a new payment agreement.
- *Cost-Effective* : Cost effective cryptographic mechanism to be applied allowing all the intermediate nodes to be able to validate the security information related to payment events in the packet.
- *Higher Security*: The system seeks to diminish all the fraudulent activities by blacklisting all the illicit users in the network.

## VI. RESEARCH METHODOLOGY

The entire proposed model is design in specific set of operations to be performed by the entities involved in the secure micropayment schema using wireless adhoc network. The IETF Simple Public Key Infrastructure Working Group is tasked with producing a certificate structure and operating procedure to meet the needs of the Internet community for trust management in as easy, simple and extensible a way as possible. The SPKI is intended to provide mechanisms to support security in a wide range of Internet applications, including IPSEC protocols, encrypted electronic mail and WWW documents, payment protocols, and any other application which will require the use of public key certificates and the ability to access them. It is intended that the Simple Public Key Infrastructure will support a range of trust models. The certificate authorization of Simple Public Key Infrastructure which combines authorization to the public key

is mechanized in the proposed system in order to combine authorization for mobile commercial payment to a user's key.

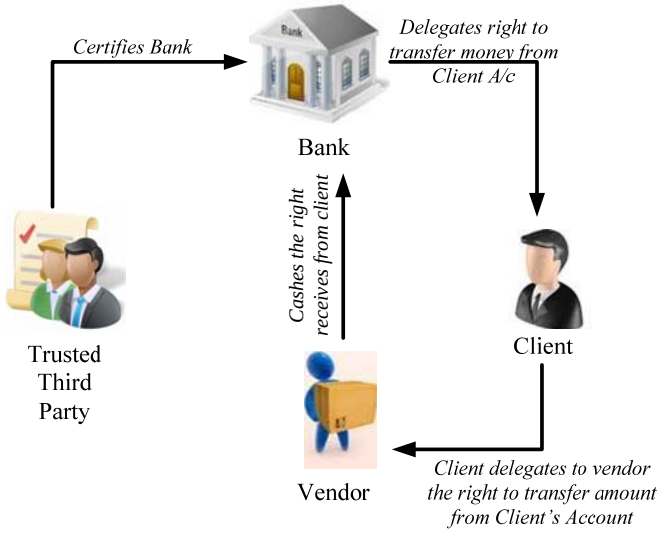



Fig.2. Payment mechanism in proposed system

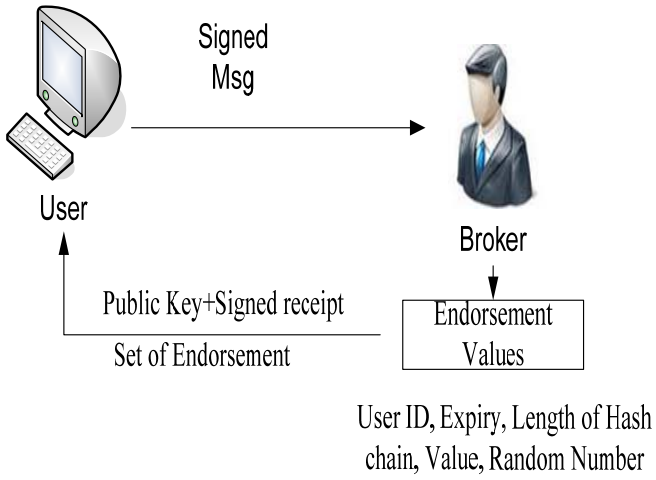



Fig.3. Broker Agreement Policy

The best feature of this model is its ability to delegate the authorization to other clients using a chain of delegates. The bank is certified by the trusted third party in the initial stage, which is done using authentication certificate for bank. In the consecutive phase, bank transmits an authorization certificate to Clients, which consist of authorization to its client C to transfer amount from client's account to the bank. The delegation flag is configured by bank which permits client to delegate this permission . Both the verification certificate along with recently designed verification certificate is transferred to client by bank. The bank identity as well as the validity of the authorization certificate of client is evaluated by client in order to check if TTP has signed the authorization certificate of bank and bank has certified that for the client. Then, client generates a new permission certificate for the vendor for transferring his rights to vendor to transfer amount from client's account. The security of the proposed system is maintained by this architecture where by implementing simple public key infrastructure by confining the rights of withdrawal. The entire cumulative certificate chain is transferred by client to vendor who analyzes its authenticity. The final stage of verification is done by bank when vendor transfer the chain to bank. The validity of the certificate is evaluated by bank to check the genuine source of the certificate (bank). After successful validation, the vendor is privileged to withdraw amount from client's account.

The authorization certificate which is frequently used consists of flag shows the validity of binding authorization and its respective delegation which is one of the prime factor of security. The proposed system assumes all the independent modules (TTP, bank, client, vendor etc) as certificate authority which is very suitable for any distributed architecture of wireless mobile adhoc network. The system reserves the chain used as it consists of confidential information related to the payment system as well as means to recognize cyber illegal users

The proposed steps are broker agreement (See Fig 3), cost and endorsement delivery, initiating payments, new route consideration, transferring tokens, and broker approval. The proposed research methodology can be explained in brief steps as following:

1. *Broker Agreement*: A broker supplies its registered and authorized user will a secure and tamper-proof token with public key pair along with highly encrypted user identity. Any micropayment schemes like credit card can be used for designing the application. The user then sends a signature message consisting of hash value and payment information which is encrypted with public key of broker. The broker generates (agreement) secret endorsement data which consists of a random number, an anchor value, length of hash chain, user-identity, and expiry of chain. These set of information is secured by private keys of broker. Therefore the broker agreement can only be deciphered by user's token. However, the security of tokens (smart cards) are not reliable as it can be deciphered, so the broker private information is appended with expiry date in order to restrict an unauthorized user in the range of mobile network to have an access on the confidential information transacted between user and broker.
2. *Cost and Endorsement Delivery*: A sender node P sends the cost request message encrypted with digital signature using their private keys to query the route of recipient node Q. All intermediate nodes attaches an certificates so that the origin node will be able to validate the digital certificates on the cost details. The data for cost reply message is returned to P. After estimating the cost involvement in routing, the encrypted broker endorsement is sent to all relay nodes in the network. These endorsements are private data, so each

user encrypt with their public key, which can be received from cost reply message. This scheme pays the intermediate routers for forwarding the packets (See Fig.4).

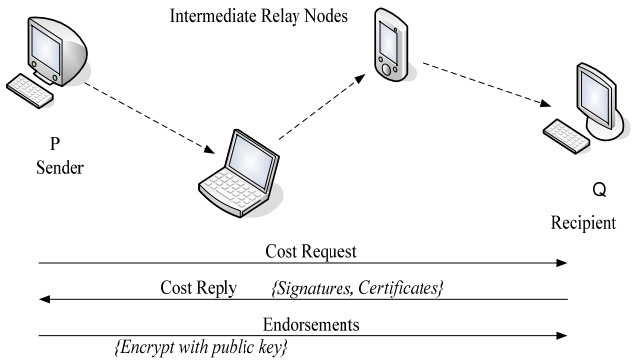



Fig.4. Endorsement-Distribution

3. *Initiating Payment*: This step is about initiating payments in the system by the user. P transmits message in his network and appends a hash token from sub-chains. The payment scheme in independent of increased used of hash values for multiple payments by the user ensuring much less network overhead. In case the intermediate relay nodes have captivated the hash values, they will not be able to decipher them without broker agreement and its respective signature.

4. *New Route Consideration*: This step is performed as wireless adhoc network quite often changes their topology dynamically. In case of new route, the system needs not to contact the any TTP. Overhead is reduced by observing the new nodes in the route and using only them for the distributing the secure endorsement.

5. *Transferring Tokens*: Here the intermediate relay node transmits the greater hash values in one chain that has spent it by the node. The user token then transmit the hash value to the consecutive broker with their endorsement digitally signed. The message and its highly encrypted contents are validated by the broker as well as issues an acknowledgement.

6. *Broker Approval*: The proposed system does support multiple brokers for reliable communication which allows any user to get associated with any broker available in the network. The user in the first network receives payment chain from the broker in that network, it assist the same user for validating the digital certificates generated by the nodes in new network when the network topology changes. The assumption to this step is that the user, broker and all the entities involved should first get themselves registered and then perform the task.

.

## VII. Architecture Design

The main motive for the highlighted methodology is to build an effective and secure e-commerce system in wireless mobile adhoc network. The proposed system highlights a very flexible architecture (See Figure 1) for secure transaction in wireless mobile adhoc network.

The architecture is basically classified into two main blocks e.g. first is PaymentEngine and second is SecurityEngine. The first block i.e. PaymentEngine basically has repository of certificates for the proposed payment schemes e.g. authorization certificates, authentication certificates, account permissions etc. The first block provides an interface for notification in direct communication with updating of repository. The maintenance user interface communicates with user. The user can be considered as innermost payment service on the machine of user. The first block i.e. PaymentEngine deploys the 2nd block i.e. SecurityEngine for signing and validating chains of certificate. The security design is accomplished by using Simple Public Key Infrastructure using cryptographic framework in java which facilitates services for signing and creating chains of certificates.

According to this architecture, bank request for a digital certificate by TTP previous to any transactions to be permitted which is quite independent from any renewal. After this bank is prepared to transfer account permission to the clients assuming all the communication is done from mobile interface in wireless adhoc network. The bank again receives its public key from client and client checks his status of permission for accessing his account assuming client has an account with bank. Exactly after the previous step is accomplished, bank generates a new permission certificate and sent it to client. Now the client is prepared to communicate with vendor for payment scheme (See Fig.5).

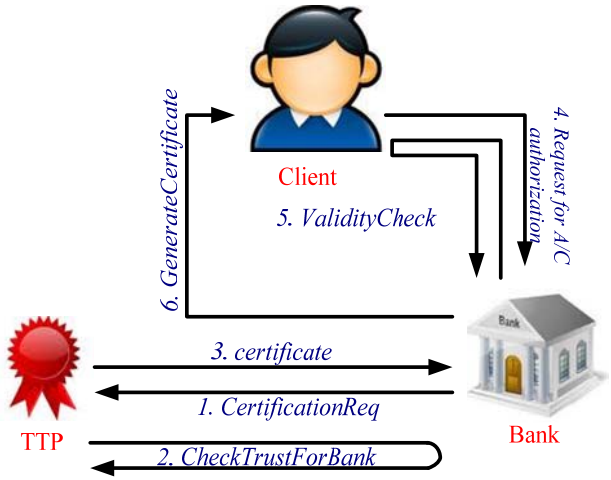



Fig.5. Permission certificate from the bank

In the second phase of the transaction (See Figure 6), when client communicates with vendor related to specific business transaction. The vendor sends a signed e-bill which includes list of TTP as there are many global TTP which user might not rely all. Client only evaluates if bank is authorized by at least one of the TTP which is conventional for vendor for secure future transaction. The communication / transaction between client and bank fails if both the party do not have certain common TTP. The transaction duration is made secure by estimating validity duration for payment. The client then

generates a deposit certificate and transfer the entire chain of certificate to vendor. The vendor accepts and evaluates the entire validity of chain. It is to be noted that for security reason, the certificate is valid for only one transaction for vendor.

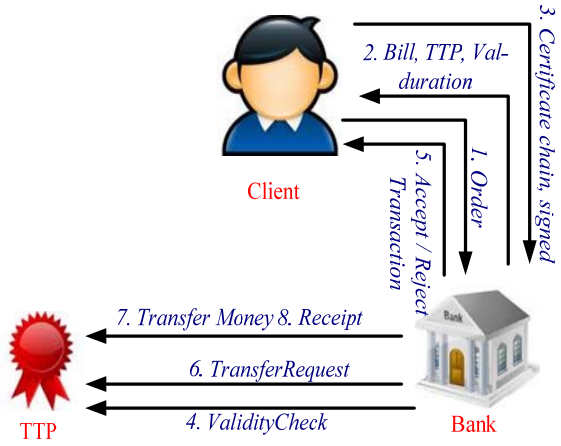



Fig.6. Payment Interaction

One of important issue with communication on offline in wireless adhoc network is that account permission for client is feasible for being invalidated without vendor module knowing about it. In order to solve this issue, the proposed system highlights account permission with very short validity duration where bank should renew certificates frequently. Therefore if the certificate has been invalidated or rejected by the bank, than it will be subjected for acceptance offline for a very shorter duration. Therefore the proposed system with short term certificates has better security in the wireless mobile adhoc network.

## VIII. Performance Analysis

The proposed system facilitates secure and reliable sets of communication with offline verification from sender to recipient node in wireless adhoc network thereby permitting a secure micropayment schemes for multiple nodes in the network by using hash functions. For providing successive endorsement distribution securely, asymmetric key are used. The proposed system is completely free from any underlying routing protocol in the wireless scenario which is very vital as routing protocols are quite dependent upon the network topologies. The long-term micropayment agreements have been eliminated because of its unsuitability in wireless adhoc network environment. Inspite of this, the cost details are securely extracted from each node in relay path to estimate a cumulative cost for forwarding the data through wireless adhoc network.

Another uniqueness in the proposed design is when a node do not have sufficient hash values for one session, then the node can be directed to some unused sub-chains by transferring a new set of endorsement. But there is a probability of loss of connection, if the system runs out of sub-chains. This phenomenon is applicable to all protocols related to micropayment system where the user registration priviledge is limited for access on the resources. The proposed system is highly favorable to the dynamic topology of wireless adhoc network as every instance the topology changes, the broker endorsement will need to be transferred to all the new nodes come across in the path. But however, it has been seen that the node mobility in such scenario as well as chain length contributes to wastage of time. However chains of higher length can be used for extreme high mobility in the network.

This section highlights the various technical requirements which are the pre-requisites for implementation of the proposed system in wireless mobile adhoc network.

### *A. General Security Issues*

The use of robust encryption along with digital signatures assures that proposed schema is not possible to illicitly decipher without specific private keys. The payment permission certificate is created only when there is a payment request and it will embed signature of both client and vendor. This is also used for identifying the dual deployment of client's payment permission certificate. The indisputability is involuntarily accomplished as all the payments are using digital signatures. The application also ensures non- traceability as flow of the transaction from one to another module can be reconstructed as the chain of certificate consists of public key of each chain. Therefore, no third person can identify the transaction information (other than bank). The propose system therefore facilities higher dimension of privacy and security.

As the centralized service consisting of revocation list will not be accessible so invalidated certificates cannot be easily cancelled in office mode. This issue is solved by using short validity duration which needs to be renewed. Therefore the entire banking application can be integrated with the mobile application very secure in wireless mobile adhoc network in offline mode. Therefore the proposed system assures pseudonymity and restricts dual payments in one session.

### *B. Serviceability*

The proposed system offers concrete usability and high dimension of creating a flexible and extremely secure system for offline e-commerce in wireless mobile adhoc network. There is no requirement of creating a new technology or abstraction from scratch for any clients to use this application. Clients has higher flexibility to make custom-build identity, delegate payment permission etc, which will assist in creating much organized e-payment system in wireless network. Moreover as the Simple Public Key Infrastructure has no dependency on operating system, so it will be highly feasible to deploy the application on any trusted handheld device like smart phone or mobile handset with OS and browser. The proposed system his highly at par with the ubiquitous application of banking system as the application do not consider a constant network infrastructure as it is designed on

wireless mobile adhoc network. Therefore, impulsive service and usage is guaranteed at any instance.

## IX. CONCLUSION

The proposed system presents a unique process oriented structural design for security scheme for micropayment which is completely independent of any trusted third party vendor. The proposed system has signified some of the instance of the non-cooperation of the nodes for providing services for micropayment system. The secured transaction adopted by the proposed system will allow the real-world micropayment system for guaranteed forwarding of the packets with highest reliability. The proposed system facilitates the routers to levy cost for each packet and also adapts to the dynamic network topology of the wireless adhoc network. The multiple routes to the recipient node with secure and encrypted cost of the packet is received by the node, depending on which appropriate direction and disseminated values of endorsement can be selected to each intermediate relay node. The intermediate node validates and initiates receiving tokens for forwarding the packets. The application concept is free from any dependency of the TTP in order to receive tokens for new route by intermediate routers. Using extra chains, it is able to initiate payments to the new node in the new network. Our future direction of research will include considering the trust and reputation management for providing more safe and more reliable operation in micropayments in wireless adhoc network. The proposed system also highlights a secure application for e-payment system in offline using wireless mobile adhoc network. The security of the application is governed by Simple Public Key Infrastructure. The formation of chains of certificate allows a distribution of the payment system by delegates. The designed model prevents dual expenditure in offline communication. The proposed system shows a flexible and robust solution for serviceability, security, and effectiveness in e-payment systems over wireless mobile adhoc network. The future enhancement work could be considered on design of security system based on specific attack on mobile adhoc network like DDoS or Wormhole attack, which is very common issue on pure mobile network deployment in larger scale of deployment.

## REFERENCE

[1] Zygmunt J. Haas, Jing Deng, Ben Liang, Panagiotis Papadimitratos, S. Sajama,"Wireless ad hoc Networks", John Wiley & Sons, Inc, 2003

[2] Yuntsai Chou, Chiwei Lee and Jianru Chung, "Understanding m-commerce payment systems through the analytic hierarchy process", Journal of Business Research, Volume 57, Issue 12, December 2004, Pages 1423-1430

[3] Neal Leavitt, "Payment Applications Make E-Commerce Mobile", IEEE Computer Society, 2010

[4] Rafael Martínez-Peláez, Francisco Rico-Novella, Cristina Satizábal and Jhon J. Padilla, "Performance Analysis of Mobile Payment Protocols over the Bluetooth Wireless Network", Whitepaper, 2008

[5] Heiko Knospe, Scarlet Schwiderski-Grosche, "Future mobile networks: ad-hoc access based on online payment with smartcards", IEEE, 2002

[6] Peter Tarasewich, Robert C. Nickerson, Merrill Warkentin, "Wireless/Mobile E-commerce: technologies, applications, and issues", Seventh Americas Conference on Information Systems, 2001

[7] R. Rivest and A. Shamir, "PayWord and MicroMint: Two Simple Micropayment Schemes", Security Protocols, LNCS 1189, M. Lomas, Ed., Springer-Verlag, 1997, pp. 69-87, http://theory.lcs.mit.edu/~rivest

[8] R. Hauser, M. Steiner, and M. Waidner, "Micro-payments based on iKP", in Proc. of the 14th Worldwide Congress on Computer and Communications Security Protection, Paris, 1996, pp.67-82, http://www.zurich.ibm.com

[9] W3C Micropayments Working Group, http://www.w3.org/ECommerce/Micropayments/

[10] D. O'Mahony, M. Peirce and H. Tewari, Electronic Payment Systems for E-Commerce, 2nd Ed., Artech House Publishers, Boston/London, 2001.

[11] L. Lamport, "Password Authentication with Insecure Communication", Communications of the ACM, vol. 24, no. 11, Nov. 1981, pp. 770-72.

[12] Nachiketh R. Potlapally, Srivaths Ravi, Anand Raghunathan, Niraj K. Jha, Analyzing the energy consumption of security protocols, Proceeding ISLPED '03 Proceedings of the international symposium on Low power electronics and design, ISBN:1-58113-682-X doi>10.1145/871506.871518, 2003

[13] C. Ellison, "SPKI Requirements", Network Working Group, Request for Comments: 2692, September 1999

[14] Zhi-Yuan Hu, Yao-Wei Liu, Xiao Hu, Jian-Hua Li, Anonymous Micropayments Authentication (AMA) in Mobile Data Network, INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies Iss: 7 March 2004,

[15] Xiaoling Dai, Oluwatomi Ayoade, and John Grundy, Off-line Micro-payment Protocol for Multiple Vendors in Mobile Commerce, Proceeding PDCAT '06 Proceedings of the Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE Computer Society Washington, 2006

[16] Min-Shiang Hwang, Pei-Chen Sung, A Study of Micro-payment Based on One-Way Hash Chain, International Journal of Network Security, Vol.2, No.2, PP.81–90, Mar. 2006

[17] Samad Kardan and Mehdi Shajari, A Lightweight Buyer's Trust Model for Micropayment Systems, WSEAS Transactions on Information Science & Applications, 2008

[18] Sung-Ming Yen, Chien-Ning Chen, Hsi-Chung Lin, Jui-Ming Wu, and Chih-Ta Lin, Improved Probabilistic Micropayment Scheme, Journal of Computers Vol.18, No.4, January 2008
[19] Lih-Chyau Wuu, Kuang-Yi Chen, Chih-Ming Lin, Off-Line Micro Payment Scheme with Dual Signature, Journal of Computers, Vol.19, No.1, April 2008
[20] Vivek Katiyar, Kamlesh Dutta, Syona Gupta, A Survey on Elliptic Curve Cryptography for Pervasive Computing Environment, International Journal of Computer Applications (0975 – 8887) Volume 11– No.10, December 2010
[21] Husna Osman, Hamish Taylor, Design of a Reputation System for M-Commerce by Ad Hoc Networking, "Design of a reputation system for m-commerce by adhoc networking," Technical Report, Dept. of Computer Science, Heriot-Watt University, 2010, pp-1-7
[22] Fouzia Mousumi, Subrun Jamil, Push Pull Services Offering SMS Based m-Banking System in Context of Bangladesh, International Arab Journal of e-Technology, Vol. 1, No. 3, January 2010
[23] Arogundade O.T, Ikotun A. Motunrayo, Olaniyi Ademola, Developing a Usage-centered e-Payment Model using Open Network System, International Journal of Computer Applications (0975 – 8887) Volume 12– No.6, December 2010
[24] Partha Pratim Ghosh, Sabyascahi Pattnaik, Gunjan Verma, Improving Existing e-payment Systems by Implementing the Concept of Cancelable Biometrics, Partha Pratim Ghosh et. al. / International Journal of Engineering Science and Technology Vol. 2(7), 2010
[25] Mohammad Al-Fayoumi, Sattar Aboud and Mustafa Al-Fayoumi, Practical E-Payment Scheme, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 7, May 2010
[26] Kaylash Chaudhary, Xiaoling Dai and John Grundy, Experiences in Developing a Micro-payment System for Peer-to-Peer Networks, International Journal of Information Technology and Web Engineering, vol. 5, no. 1, 2010
[27] Charles K. Ayo, Wilfred Isioma Ukpere, Design of a secure unified e-payment system in Nigeria: A case study, African Journal of Business Management Vol. 4(9), pp. 1753-1760, 4 August, 2010
[28] Jian-Sen Wang, Fuw-Yi Yang, and Incheon Paik, A Novel E-cash Payment Protocol Using Trapdoor Hash Function on Smart Mobile Devices, IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.6, June 2011
[29] D. Chaum, "Blind signature for untraceable payments", In: Proceedings of advances in Cryptology, Springer-Verlag, New York, pp.199-203, 1983.
[30] W. S. Juang and H. T. Liaw, "A practical anonymous multi-authority e-cash scheme", Applied Mathematics and Computation, Vol. 147, No. 3, pp. 699-711, 2004.
[31] Y. Y. Chen, J. K. Jan, and C. L. Chen, "A novel proxy deposit protocol for e-cash systems", Applied Mathematics and Computation, Vol. 163, No. 2, pp. 869-877, 2005.
[32] C. L. Chen and M. H. Liu, "A traceable E-cash transfer system against blackmail via subliminal channel", Electronic Commerce Research and Applications, Vol. 8, No. 6, pp. 327-333, 2009.
[33] C. C. Chang and Y. P. Lai, "A flexible Date-attachment Scheme on E-cash", Computers & Security, Vol. 22, No. 2, pp.160-166, 2003.
[34] W. S. Juang, "D-cash: A flexible pre-paid e-cash scheme for date-attachment", Electronic Commerce Research and Applications, Vol. 6, No. 1, pp. 74-80, 2007.
[35] C. C. Chang, S. C. Chang, and J. S. Lee, "An on-line electronic check system with mutual authentication", Computers & Electrical Engineering, Vol. 35, No. 5, pp. 757-763, 2009.
[36] W. K. Chen, "Efficient on-line electronic checks", Applied Mathematics and Computation, Vol. 162, No. 3, pp. 1259-1263, 2005.
[37] J. E. Hsien, C. C. Hsueh, and C. Y. Chen, "An electronic traveler's check system", Conference on Theory and Practice for Electronic Commerce, pp. 164–169, 2001.
[38] H. T. Liaw, J. F. Lin, and W. C. Wu, "A new electronic traveler's check scheme based on one-way hash function", Electronic Commerce Research and Applications, Vol. 6, No. 4, pp. 499-508, 2007.
[39] Kevin Curran, Shane Dempsey, "Enhancing Bluetooth security for m-commerce transactions", Adv. Engg. Info., 2011
[40] Jian-Sen Wang, Fuw-Yi Yang, and Incheon Paik, "A Novel E-cash Payment Protocol Using Trapdoor Hash Function on Smart Mobile Devices", IJCSNS International Journal of Computer Science and Network Security, Vol.11 No.6, June 2011
[41] Natarajan Vijayarangan, "A system and design of Extensible Authentication Protocols based on ECC and SKE mechanisms for mobile and wireless communications", Advances in E-Activities, Information Security and Privacy, 2011
[42] Saurabh Panjwani, Prasad Naldurg, Raghav Bhaskar, "Analysis of Two Token-Based Authentication Schemes for Mobile Banking", Technical Report of Microsoft Research, 2010
[43] Laetitia Chaix and Dominique Torre, "Different models for mobile payments, research paper, 2010
[44] Haifeng Wu, Xuan Li, Weihui Dai, Weidong Zhao, "Mobile Payment Framework Based on 3G Network, Proceedings of the Third International Symposium on Electronic Commerce and Security Workshops(ISECS '10) Guangzhou, P. R. China, 29-31, pp. 172-175, July 2010,
[45] Mobile Payments, A White Paper by Microsoft and M-Com, 2011
[46] http://en.wikipedia.org/wiki/Mobile_payment [Accessed on 4th-Aug, 2011]
[47] http://en.wikipedia.org/wiki/Micropayment [Accessed on 30th July, 2011]

[48] Xiaoling Dai1 , John Grundy and BruceWN Lo, Comparing and contrasting micro-payment models for E-commerce systems, Info-tech and Info-net, Proceedings. ICII 2001 - Beijing. International Conferences, 2001

# Semantic Probabilistic Modelling of novel routing Protocol with Implication of Cumulative Routing Attack in Mobile adhoc network

Anil G. N.
*Asst. Prof*: Dept of CSE.
BMS Institute of Technology
Bangalore, India

Dr. A. Venugopal Reddy
*Prof*. & Principal,
University College of Engineering, Osmania University
Hyderabad, India

***Abstract*— The proposed system presents a novel approach for modelling along with mitigating various types of routing attacks in mobile adhoc network considering AODV protocol. Majority of the previous research work are either explored differently for security or routing protocols. The system identifies the susceptibility of the routing attack over the dynamic topology of mobile adhoc network where it has assumed a faster propagation of the infection towards the nodes. To make the analysis more challenging, the protocol also designs a sophisticated adversary module which is resilient against any types of preventive measure being adopted. The proposed system therefore used probabilistic approach for modelling the routing attack scenario over MANET. The uniqueness is that majority of the prior research work has focused on one type of routing attack, whereas the proposed system is experimentally evaluated for cumulative routing attack. The simulation results show highly contrastive result when compared with frequently used current algorithm for mitigating routing attacks.**

***Keywords-Routing Attack, Mobile Adhoc Network, Security, AODV, Probabilistic approach***

## I. INTRODUCTION

Mobile Adhoc Network consists of independent wireless mobile nodes which group together to form a momentary wireless network without any assistance of any centralized management or fixed infrastructure. Internet Engineering Task Force (IETF) has MANET working group (WG) that is devoted for developing IP routing protocols [1]. One of the huge concerns in mobile adhoc network is the maintaining efficient, robust, and secure routing protocols, which has attracted many researchers. Routing protocols are normally required for maintaining efficient transmission among the mobile nodes by exploring the network topology, which in this case is always dynamic. It also designs a route for pushing the data packets and also manages the routes among the pair of mobile nodes. One of the fundamental problem with majority of the routing protocol is that the routing protocol relies on all the mobile nodes present in the network and depending on the situation that these mobile nodes will perform or collaborate appropriately; but there is a higher feasibility of circumstances where certain specific set of nodes may not behave appropriately giving rise to suspicious factor. Unfortunately, majority of the routing protocols in mobile adhoc network is witnessed for declined performance at the time of communicating with large scale of misbehaving nodes, which definitely sustains the course of route exploration but also disrupt the course of data rendering the routing protocol to resume again the route exploration procedure or to chose an unconventional route in case it is available. Moreover the newly opted route has the feasibility of possessing a few malicious nodes, resulting in failure of new route too. Such methodology iterates till the sender node confirms that the data will not be able to transmit ahead. The fundamental issue with frequently used routing protocols is that they rely all mobile nodes of network and based on the assumption that nodes will behave or cooperate properly but there might be a condition where some nodes are not behaving properly. Majority of the adhoc network routing protocols becomes inefficient and shows reduced performance while mitigating with big number of misbehaving nodes. Such set of misbehaving nodes support the flow of route discovery traffic but interrupt the data flow, causing the routing protocol to restart the route-discovery process or to select an alternative route if one is available.

Mobile adhoc network should posses a better and effective security as various upcoming application based on MANET are on its way in future. Mitigating the routing issues will create a better, efficient, and secure application in mobile adhoc network. Various types of attacks in mobile adhoc network like Wormhole attack, Black hole attack, Sybil attack, flooding attack, routing table overflow attack, Denial of Service (DoS), selfish node misbehaving, impersonation attack etc. has to be researched on more for better result. The proposed paper will present a framework for mitigating majority of the types of routing attack using probabilistic approach. The proposed system has large dimension of testing conducted to check the efficiency of routing protocol using AODV on majority of routing attack in mobile adhoc network.

In Section II, we will discuss about the previous research work in this area followed by Section III about various categories of routing attack. Section IV highlights proposed system followed by implementation in section V. In depth discussion of research performance analysis is done in section-VI followed by conclusion in section-VII.

## II. RELATED WORK

Recently, numerous approaches have been proposed to deal with the node non-cooperation problem in wireless networks. They generally can be classified into two main categories: reputation systems and price-based systems. We use a monitoring and reputation system [2] as the basic setting for regular nodes. Many related works also use reputation systems [3]–[5] and a game theory model [6] to analyze the problem. Some recent works have studied the incentives for malicious nodes and modeled their behavior more rationally. In [7], Liu et al. present a general incentive-based method to model the attackers' intents, objectives, and strategies. In [8], Theodorakopoulos and Baras further study the payoff of the malicious nodes and identify the influence of the network topology. However, the good nodes' behavior in [9] is simple, and it fails to consider the possibility that an attacker might choose different attack frequencies toward different opponents.

The security problem and the misbehaviour problem of wireless networks including MANETs have been studied by many researchers e.g. [9], [10], [11], [12].Various techniques have been proposed to prevent selfishness in MANETs. These schemes can be broadly classified into two categories: credit-based schemes and reputation based schemes. The basic idea of credit-based schemes is to provide incentives for nodes to faithfully perform networking functions. Nodes get paid for providing services to other nodes. When they request other nodes to help them for packet forwarding, they use the same payment system to pay for such services.

Sanjeev Rana [13] has created a mechanism with the help of which it prevents various replay attacks and also activate the neighboring nodes to control the behavior of its neighbors to thwart active attacks. Rakesh Kumar e.t. al [14] has implemented a prototype of key management service by using genetic algorithm. Rajib Das e.t. al [15] has proposed a solution against black hole attack and has illustrated the effect of black hole attack on network performance. However, the results cannot be considered as optimal. Kannan e.t. al [16] has published an extensive survey on various attacks, and their respective countermeasures with respect to vulnerability in routing protocols. Aishwarya Sagar [17] has proposed an approach based on reputation system that deals with routing misbehaviour and consists of identification and separation of misbehaving nodes. Hariharan e.t. al [18] has proposed a new technique termed as recommendation based on identification of routes with misbehaved nodes. Usman e.t. al [19] have analyzed the effects of different types of jammers using Conservation of Flow (CoF), which has been useful for detecting other attacks, in the wired networks. Abbas [20] have categorized reputation based schemes based on monitoring approaches: active and passive based acknowledgments. Finally, the authors have discussed their pros and cons as well as some other important identity related issues.

Depending of the patterns of the intrusion, attacks towards mobile adhoc network can be categorized into active or passive attack. Not only this, the attacks can be also further classified into internal or external attack. In association with the victim node, the attack can be again classified into routing packet or data packet attacks. In case of routing packet attack, the malicious node resist existing routes from being utilized and also it spoofs other non-existing routes for alluring data packets to be forwarded to them.

Although there are number of research conducted in past [21],[22],[23], [24], [25] for analyzing routing attacks on mobile adhoc network. Important routing attacks are fabrication, blackhole, and alteration of various fields in routing packets e.g. RREQ, RREP, RERR message, etc. Research work conducted in [26], [27], [28] discusses about some mitigating techniques for safeguarding the routing protocols in mobile adhoc network. Although these set of research work can successfully resist illegitimate nodes from participating the network, but unfortunately, it was found to increase the significant network overhead with respect to key exchange as well as authentication with restricted intrusion eradication.

The resistance based approach are also found less efficient for mitigation from malicious intruders who already have the confidential information for rendering communication by themselves in the mobile adhoc network. The prior research work has also seen the introduction of Intrusion Detection System for mobile adhoc network. Unfortunately, due to the dynamic topology of mobile adhoc network, majority of such research work are modeled to be scattered and possesses cooperative data-structure.

Specification-based approaches, for example DEMEM [29], C. Tseng et al. [30] and M. Wang et al. [31], monitor network activities and compare them with known attack features, which are impractical to cope with new attacks. A completely new work done in same field called as Intrusion Response System in mobile adhoc network has being discussed in [32] which detaches the malicious node, once identified, depending on their reputation system. Unfortunately, the work fails to be at par with efficient IDS system. The Table.1.will highlight specifically all the prominent research work being done towards securing routing protocol in mobile adhoc network.

Table.1. Prior research work

| Year | Authors | Problem Focused | Approaches used | Results Obtained |
|---|---|---|---|---|
| 2009 [33] | M.K. Jeya Kumar R.S. Rajesh | Cumulative routing Issues | Designed a mobility model using Random waypoint | AODV performs better than other routing protocols. |
| 2009 [34] | Abdul Rahman Zuriati Zukarnain | Link breakage | Designed a mobility model using Random waypoint | AODV performs better than other routing protocols. |
| 2009 [22] | Nishu Garg R.P.Mahapatra | Performance degradation due to routing issues | Just discussed about security consideration for effective routing | Not optimized result |
| 2009 [35] | Dipankar Deb Srijita Barman Roy Nabendu Chaki | GPS-free positioning systems | Designed Location Aided Cluster Based Energy-efficient Routing | Lowering mean hop and hence in utilizing the limited energy of mobile nodes. |
| 2009 [36] | E.A.Mary Anita V.Vasudevan | Black hole attack | Designed Security in Multicast Ad-hoc On Demand Distance Vector | Better result for Black Hole attack only |
| 2009 [37] | Ashwani Kush P. Gupta C.Jinshong. Hwang | Security in Routing protocol | Designed a Power Aware Virtual Node Routing Protocol | Not optimized result<br>Increases Network Overhead |
| 2009 [38] | Sheenu Sharma Roopam Gupta | Black hole attack | measuring the packet loss in the network with and without a blackhole | Only 26% reduction in network performance in presence of Blackhole attack |
| 2009 [39] | Cong Hoan Vu, Adeyinka Soneye | Collaborative Black hole Attacks | Designed a simulation to check the performance | Only resistive against Blackhole attack. |
| 2010 [40] | Irshad Ullah Shoaib ur rehman | Black hole attack | Studying Blackhole attack on OLSR and AODV | Is not effective on DSR, TORA, GRP etc. |
| 2010 [41] | Shishir K. Shandilya Sunita Sahu | RREQ Flooding Attack | Designed a distributed cooperative model in which all the node locally run the intrusion detection code and cooperate with each other to detect and prevent flooding attack in the network. | Results completely dependent on threshold value.<br>The proposed result delays the detection of misbehaving node |
| 2010 [42] | Akanksha Saini Harish Kumar | Effect Of Black Hole Attack On AODV | Designed a simulation to check the performance | The experiment didn't reached the better results for ensuring protection from blackhole attack on AODV routing protocol |
| 2010 [43] | Aishwarya Sagar Anand Ukey Meenu Chawla | Packet Dropping Attack Routing Misbehavior | Designed a simulation to check the performance | Results doesn't guarantee that ACK packets are genuine and no work done in punishing misbehaving nodes. |
| 2010 [44] | Moitreyee Dasgupta Choudhury Chaki | Routing Misbehavior Impact of rushing attack implemented by malicious nodes (MNs) on AODV routing protocol | Designed RREQ forwarding mechanism | Better result for Rushing attack only however. |
| 2011 [45] | Kannan Maragatham | Study of various attack | Just a theoretical Paper | -N/A |
| 2011 [46] | Amrit Suman | Work hole attack | analyze three ad-hoc routing protocols AODV, DYMO, FISHEYE against wormhole attack in wireless network. | Better result for worm hole attack |

## III. ROUTING ATTACKS

The prominent job of the routing protocol is to explore the topology in order to ensure that every node can get the access on current map of the network for designing routes in its destination. The routing attack can be represented as shown in Fig.1. where a malicious node (MN) can completely absorb the network traffic by introducing themselves within the network link of sender node to recipient node along with intermediate nodes (IN-1, IN-2) and thereby possessing the unauthorized control over the mobile adhoc network.

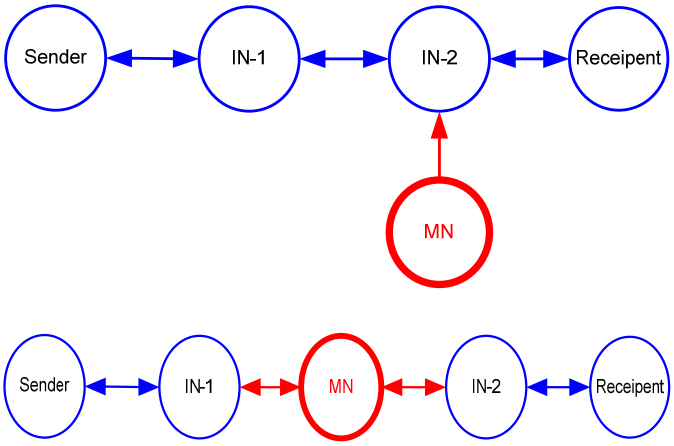



Fig.1. Representation of Routing Attack

The theoretical understanding of the attacks in network layer can be classified as routing attack and packet forwarding attack [47]. The concept of routing attack will be related to performing any activity for broadcasting updates on routing which do not trail the configuration of the specified routing protocols where the particular attack pattern is associated to the routing algorithm deployed by mobile adhoc network. The infected or malicious routing nodes can initiate an attack in mobile adhoc network by deploying different ways. Some of the major type of the routing attack as follows:

1. *Routing Table Overflow Attack*: It normally takes place in proactive routing algorithm. The main purpose of this type of attack is to create big set of routes so that designing of new routes can be resisted.

2. *Routing Table Poisoning*: in this type of attack, the malicious node forwards fake routing updates resulting in routing table poisoning.

3. *Packet Replication*: Here the malicious node duplicates the out of date packets for consuming bandwidth and unwanted resource consumption.

4. *Rushing Attack*: Such attack is highly targeted on on-demand routing protocol.

5. *Route Cache Poisoning Attack*: It uses the benefit of promiscuous mode of updating routing table, which normally occurs when confidential information within the routing table is either modified, erased, or maliciously written with fake information.

**Attacks on Particular Routing Protocol**

Various attack has been reported in the prior research work which is very much specific to the commonly used routing protocol in mobile adhoc network which only surfaces due to designing services of routing without assuming prime issues in security. The routing protocols specific to the attack are as follows:

1. *AODV*: The adhoc on-demand distance vector is reactive routing algorithm, where the attacker [22] all attempt to broadcast a route with less distance parameter than original or broadcast a fake routing information to disrupt the routing.

2. *DSR*: Dynamic Source Routing which is almost similar to AODV, where it is highly feasible to alter the source route listed in the control message (RREQ, RREP) by the attacker.

3. *ARAN*: Authenticated Routing for Adhoc Network is also a type of on-demand routing protocol. Although, ARAN has some best features for security in adhoc network, but still it cannot stand for rushing attack.

4. *ARIADNE*: It is also an on-demand secure routing protocol which is based on dynamic source routing. Although ARIADNE is robust for denial of service attack, but still it cannot mitigate wormhole and rushing attack [48].

5. *SEAD*: It is designed on Destination Sequence Distance vector. It can encounter against replay attack using cost effective cryptographic algorithm, but it cannot stand against wormhole attack [48].

Other types of attacks commonly found in literature are Wormhole Attack, Blackhole Attack, Byzantine Attack, Rushing Attack, Resource Consumption Attack, and Location Disclosure Attack

**Countermeasures for Routing Attacks**

The vulnerability of the network layer is very highly prone for routing attack in comparison to other layers in mobile adhoc network, which induces a diversified threats in security. Deploying algorithms in security considering routing protocols will only facilitate the countermeasures against such lethal attack which is very difficult to identify. Source authentication as well as message integrity technique can be

used for non-passive attack e.g. modification of routing information content. Use of message authentication code, digital signature, hashed MAC as well as one-way hash MAC key chain can be deployed for such mitigation technique. While wormhole attack can be mitigated by using unchangeable and self-sufficient physical parameter e.g. delay in time and geographical position. The work done in [49] already proved adoption of packet leashes for mitigating such issues. Another most frequently used technique is IPSec on network layer in world wide web to facilitate definite layer of privacy and security. ARAN is another efficient routing protocol which provides protection from various types of attacks e.g. corruption of hop counts, altering of sequence number, IP spoofing, DDoS, fabrication of source route [50]. Finally, work done in [51] also highlights use of security technique to mitigate blackhole attack by paralyzing the ability of reply of an intermediate mobile node, so that destination node never receives the reply message.

## IV. PROPOSED SYSTEM

The proposed system presents a framework for contrastive analyzation of routing protocols where the routing attacks can be determined. Majority of the prior research work has focused on building either a mathematical model or any analytical model considering one of the type of routing attack in mobile adhoc network. The problem with such approach is that it can better thwart for one of the routing attack while become inefficient for other types of routing attack. So, due to this research gap, the proposed system has focused on designing a hybrid framework which can model almost all types of routing attack in mobile adhoc network thereby acting as an effective solution for identifying the sectors of routes which are compromised or about to be compromised.

The proposed system can be classified into following modules e.g. network model, cryptographic model, attacker model.

### A. *Network Model*:

The current work of mechanizing the security in routing protocol is designed considering group of nodes N and routes R, which can be represented mathematically as G={N, R}as directed graph. The route R is completely dependent on factors like current position of node, relationship, and charecteristics of the mobile nodes, medium of communication, and MAC layer. The dispatcher and destined nodes can be depicted as D and d, which is constructed depending on decision taken by routing protocol. One or multiple routes will be designed considering set of sequential R for a given set of dispatcher node D and destined node d. Cumulative route $CR_{D,d}$ is designed for all the links considered from D to d. Let $F_t$ signifies the part of the travel from D to d such that it travels the path $t \in CR_{Dd}$. The cumulative route $CR_{Dd}$ can be depicted as route sub-graph $G_{Dd}$ of G possessing mobile nodes and directed graph travelled by atleast one of the routes $t \in CR_{Dd}$. The routing protocol using AODV is designed based on segregating the spatial factors depending on packets forwarded along the diversified routes. The consideration is made for both single and multi-paths.

### B. *Cryptographic Model*:

The module will be responsible for maintaining security of the packets by assigning cryptographic keys. The model considers $S_{key}$ as group of symmetric security keys and $P_{key}$ be equivalent group of public keys. If i be node number considered than $i \in N$, which is allocated with $S'_{key}$ such that $S'_{key} \subseteq S_{key}$ and also public tag substitution key $P'_{key} \subseteq P_{key.}$ The common set of the keys shared among i and j as $S_{key(i,j)} = S_{key(i)} \cap S_{key(j)}$, which is the criteria for permitting transmission of packets between i and j when $S_{key(i,j)} \neq 0$. The representation is as shown in Fig.2. It is also considered that the model will use $S_{key(i,j)}$ shared keys completely in order to protect the specified route (i, j). Therefore, the proposed model should have some common keys in $S_{key(i,j)}$ for secure communication in specified route. Not only this, the model will also consider the computation of $P_{key(i,j)}$ as $P_{key}(i) \cap P_{key(j)}$ for the purpose of estimating the group of shared key $S_{key(i,j)}$.

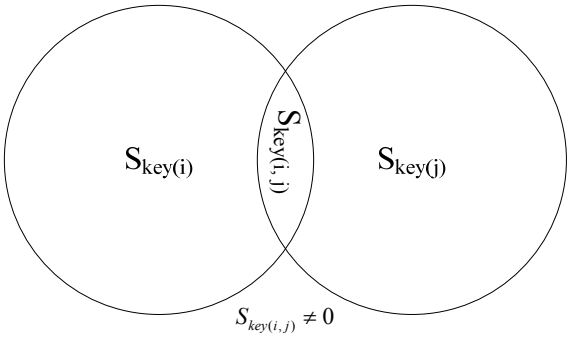



Fig.2. Set Representation of Cryptographic assumption

The category of public tag substitution design will posses any rule which facilitates required data from any other mobile nodes $j \in N$ in order to evaluate group of $P_{key.}$ as public broadcast mechanism. The security of the design of the routing protocol starts from this phase a rule is designed to provide dual layer of security for any message being communicated. This module is created to show that our attacker module is a stronger module to decrypt even this security layer, thereby assist the framework to catch hold of the attacker by determining the infected routes till that instant. This dual layer of security will facilitate data for node j for only estimating $P_{key(i,j)}$ with respect to node i without furnishing any information to other node j.

### C. *Attacker Model*:

Our previous work has already focused on the modelling the behaviour of the attacker node for preventing the decamping mechanism. The uniqueness part of the proposed system is the design of this attacker module where we are considering that

this module is extremely strong enough to decrypt any of the information transacted between any authorized mobile nodes by invoking any types of routing attack. The main intention of this module is to intrude or initiate any of the routing attack along with infection spread from dispatcher node D to destined node d with minimum cost of attack. But this time the attacked module is enhance with additional capability by which they can attach an unit cost in resource expenditure needed to initiate an attack. As in mobile adhoc network, there is no digital certificate authentication among the nodes so, we also consider that this module will attempt to give rise to all issues like route disruption, node isolation, and resource consumption in the defined scenario of the mobile adhoc network and they perform all this by extracting the secure keys from the authorized nodes. The model also assumed to posses all the route information $G_{Dd}$ using our previous model.

## V. IMPLEMENTATION

The proposed system is implemented on 32-bit windows OS with 1.84 GHz dual core processor using Java Platform. The framework of the project work is has basically two stages of operation for the users. The first stage enables the user to select and configure their network choosing from key analysis, link analysis, as well as route analysis. The key analysis section is further classified into 5 different types: basic, polynomial, hybrid, broadcast, and public type. Once after proper selection of the key analysis mode, the key development mode gets surfaces in the framework, which is further divided into random (or binomial) and pulse type. Link Analysis mode operates in the same way. Route Analysis is activated along with the options of Traffic type, flow deployment, and routing type. The user can select the traffic type with an option of data collection, data distribution, and peer activity. The next option of flow deployment is provided with route to closest sink and route to random target. The last option of routing type is classified for the user into three types e.g. multipath, dependent, and end-to-end type. Once the selections is set, the action mode is activated to display the network in the simulation mode. To construct the proper network, the user has to feed the proper parameters related to the node information, which includes number of nodes, its respective deployment area, radio range, connectivity, key ring size, maximum hop cost and distance, and multipath spread per hops. Insertion of the proper parameters will check for neighborhood size, maximum connectivity, and key pool, based on which the network will be constructed. The simulated network will show the traffic, link, potential routes, key information, initiate attacks. Finally, the user can visualize the compromised values related to key, nodes, links, traffic, and route.

The proposed model also designs a route sensitivity parameter (RSP) in order to compute optimal security standard for the transmission being active on specified route $CR_{Dd}$. Majority of the prior research work towards security of routing in mobile adhoc network is more focused on identification of the attack only after the event of attack has already being bypassed. Such approaches are beneficial for only understanding the flaws in design of routing protocol. Unfortunately, the routing attacks on dynamic MANET scenario are very much latent within the wireless adhoc network and their propagation model is almost impossible to predict. Therefore the proposed model will use probabilistic approach for generating a various diversified routes for any given application of mobile adhoc network and visualize the effectiveness of routes by invoking attack in the routes to estimate the safe routes and unsafe routes. For the effectiveness of the result, the model will implement Greedy Heuristic Algorithm.

**Algorithm**: Probabilistic Model of identifying routing attack in given MANET scenario

**Input**: Node parameters

**Output**: Identification of infected routes

**Steps**:

1 ***Initialize*** *mobile node parameters*
2 *Define key types*
3 *{public, broadcast, polynomial, hybrid}*
4 *Define route types*
5 *{multipath, end-to-end}*
6 ***Initialize*** *maximum hop distance*
7 *Estimate neighborhood size*
8. $Size_{(neigh)} = \{(No.\ of\ Nodes)\ .(\pi).(Radio\ Range)^2\}/\ Deployment\text{-}Area$
9 *Design network module*
10 *Design Cryptographic module*
11 *Design attacker module*
12 ***Switch*** *Case (Route Susceptibility):*
13 $S_{key(i,j)} \neq 0$
14 $S_{Dd}(\phi) = 0$
15 $S_{Dd}(A_{nodes}) = 1$
16 $0 < S_{Dd}(A_{nodes}) < 1$
17 *Estimate current values of all parameters*
18 ***If*** *Cost in reduced and* $S_{Dd}(A_{nodes}) = 1$
19 *Estimate Anodes and* $CR_{Dd}$.
20 *Estimate Cost*

$$\sum_{i \in A_{nodes}} Cost_i$$

21 ***Else***
22 ***Go*** *to Step (18)*
23 *Estimate probabilistic infected routes*

$$\sum_{i \in N} \frac{S_i(A_{nodes})}{Cost_i}$$

24 ***End***

The proposed model will estimate the impact of routing attack on the designed security routing protocol considering specified cumulative route $CR_{Dd}$ with the initiation of attack on group of nodes $A_{nodes} \subseteq N$. Let us consider $S_{key(comp)}$ as group of keys being corrupted by the attacker module, which will mean that any packets transmitted through $CR_{Dd}$ which was already encrypted with $S_{key(i)}$ or $S_{key(j)}$ will definitely get compromised by the malicious nodes present within that route. The considered route (i, j) or (D, d) $\epsilon$ $P_{key}$ is attacked if and only if

$S_{key(i,\ j)} \subseteq S_{key(comp)}$ and let $P_{key(comp)}$ represents all the attacked routes. Therefore the design of attack on complete route from dispatcher node D to destined node d will represent that any message being communicated using the specified route will definitely get corrupted by the $A_{node}$. Not only this, the design of the proposed routing protocol also considers the route susceptibility for routing attack when it comes under any of the following criteria:

- $S_{Dd}(\phi) = 0$, which means there is no attack if there is no routes from Dispatcher node D to destined node d.
- $S_{Dd}(A_{nodes}) = 1$ , which means that $CR_{Dd}$ is only attacked when there is presence of atleast 1 $A_{nodes}$.
- $0 < S_{Dd}(A_{nodes}) < 1$, which means the maximum and minimum intensity of attack considering complete route is not attacked but only a portion of it is infected due to routing attack.

## VI. Performance Analysis

The simulation is performed for 200 mobile nodes using random distribution in the simulation area. The model is designed considering the arbitrary allocation of 45 keys. The proposed model is evaluated for its efficiency considering comparative analysis with the prior research work conducted in security of routing protocols in mobile adhoc network. The frequently used approaches are Genetic Algorithm [52], Neural Network [53][54], Artificial Immune System [55][56], and Classification Algorithm [57] using Support Vector Machine (SVM).

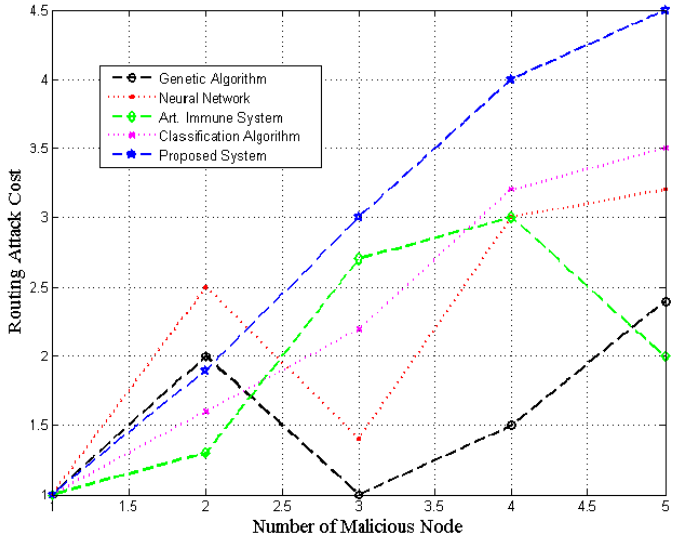



Fig.3. Independent Routing Attack

The Fig.3. shows the performance analysis when conducted for independent route. The bottom line is mobile nodes are arbitrarily attacked independently causing the aggravation of malicious nodes to initiate routing attack. However, the proposed system has higher detection rate as compared to prior research work shown.

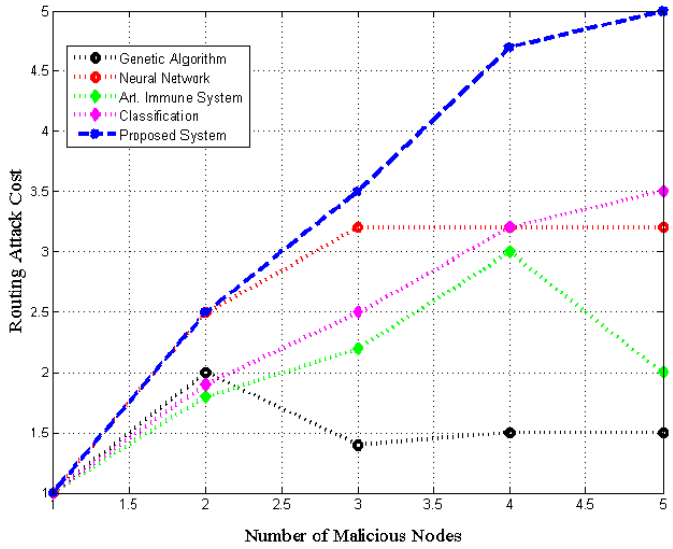



Fig.4.Intrusion in Privacy in routing attack

Fig.4. shows the performance analysis for intrusion in privacy policy maintained at each nodes. As the routing attack has iterative and sequential propagation model, so quantity of the infected routes are maximized in terms of cost. It can also be seen that by introducing the proposed protocol, the performance of attacker for initiating routing attack is reduced by maximizing the improbability in route susceptibility.

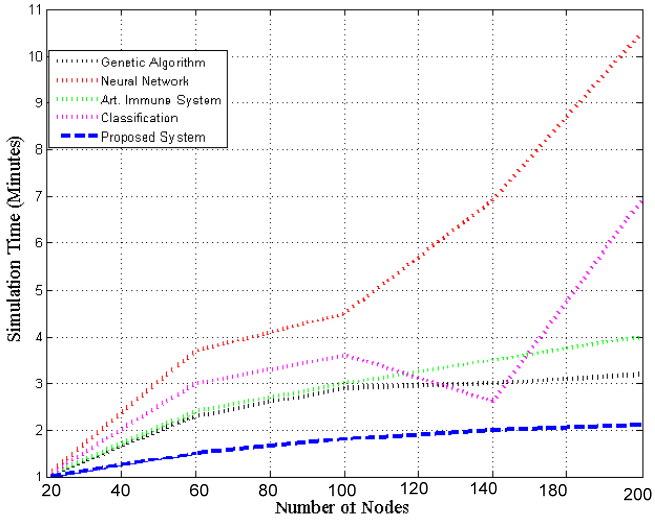



Fig.5. Simulation Speed comparison

The efficiency of the proposed algorithm is tested by observing the simulation speed required for identifying the routing attack in mobile adhoc network along with increase of number of malicious nodes at the run time of simulation as shown in Fig.5. Already consideration of dynamic topology of mobile adhoc network posses issues in the design and implementation of the algorithm, but the challenge portion of the performance analysis is made more sophisticated by

introducing more number of user defined multiple attack mobile nodes ($A_{nodes}$) at the run time of the simulation. This experiment is done to check the efficiency of the proposed algorithm to identify many attack variables which is not even programmed. The simulation result in Fig.5. clearly shows that proposed system takes comparatively less time. The graphical analysis also shows highest peak for neural network approach due to inclusion of learning phase of the algorithm, which consumes enough time for performing simulation. This fact should be kept in mind as propagation of the routing attack is very faster which starts infecting even in a matter of seconds depending upon the existing security loophole factor existing in the wireless network. It can be clearly seen that the proposed algorithm has better contrastive result in comparison to most frequently used algorithms used in current research for analyzing the security issues in routing protocol in mobile adhoc network. The implementation of the proposed system facilitates the better visualization for route susceptibility; however, an efficient route susceptibility parameter can be designed with slight alteration. The design also guarantees if any compromised route is considered for analyzing routing attack by replacing $CR_{Dd}$ by cost estimation in the similar route including direct route considering single hop type (D, d). The routing attack on unit route vector $t \in CR_{Dd}$ is more than enough for permitting the attacker to recuperate a portion $F_t$ of the route from D to d.

## VII. CONCLUSION

The proposed paper has examined the issues in designing new efficient and secure routing protocol considering all the routing attack susceptibility parameter in order to enhance the efficiency of the proposed protocol using AODV. A mathematical model is design with algorithm for estimating the impact of majority of the routing attack on mobile adhoc network using probabilistic approach using greedy heuristic algorithm. A sophisticated attacker module is design which can initiate a routing attack in our case in order to understand whether the proposed algorithm can efficiently trace the infected routes. Majority of the implementation done by enhancing cryptographic approach is considered to increase the network overhead which results in poor performance in the network. But our algorithm executes in less than 2 minutes to simulate a large scale scenario of routing attack in user-defined consideration of mobile adhoc network. Therefore, we have evaluated the simulation speed of the proposed design by adding up multiple malicious nodes at the runtime of the simulation. A comparative analysis is performed with proposed system against most frequently used algorithm like Genetic Algorithm, Neural Network, Artificial Immune System, Classification Algorithm using SVM to see that our system has contrastive result in comparison.

.

## REFERENCE

[1] http://www.ietf.org/dyn/wg/charter/manet-charter. Accessed on 6th Dec, 2011
[2] S. Buchegger and J. Boudec, “A robust reputation system for P2P and mobile ad-hoc networks,” in Proc. 2nd Workshop Econ. Peer-to-Peer Syst., 2004, pp. 403–410.
[3] F. Li and J. Wu, “Mobility reduces uncertainty in MANETs,” in Proc. IEEE INFOCOM, 2007, pp. 1946–1954.
[4] A. Josang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision,” Decis. Support Syst., vol. 43, no. 2, pp. 618–644, Mar. 2007.
[5] F. Li, A. Srinivasan, M. Lu, and J. Wu, “Uncertainty mitigation for utility-oriented routing in MANETs,” in Proc. IEEE GLOBECOM, 2007, pp. 427–431.
[6] F. Li and J. Wu, “Hit and run: A Bayesian game between malicious and regular nodes in mobile networks,” in Proc. IEEE SECON, 2008, pp. 432–440.
[7] P. Liu, W. Zang, and M. Yu, “Incentive-based modeling and inference of attacker intent, objectives and strategies,” ACM Trans. Inf. Syst. Secur., vol. 8, no. 1, pp. 78–118, Feb. 2005.
[8] G. Theodorakopoulos and J. Baras, “Malicious users in unstructured networks,” in Proc. IEEE INFOCOM, 2007, pp. 884–891.
[9] Kejun Liu, Jing Deng, Pramod K. Varshney, and Kashyap Balakrishnan, “An Acknowledgment-Based Approach for the Detection of Routing Misbehaviour in MANETs”, IEEE Transactions on mobile computing, Vol. 6, NO. 5, May 2007.
[10] Dhanalakshmi, Dr.M.Rajaram ,” A Reliable and Secure Framework for Detection and Isolation of Malicious Nodes in MANET”, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.10, Oct2008
[11] Zan Kai Chong, Moh Lim Sim, Hong Tat Ewe, and Su Wei Tan ,” Separation of Detection Authorities (SDA) Approach for Misbehavior Detection in Wireless Ad Hoc Network”,PIERS Online, VOL. 4, NO. 8, 2008.
[12] S. Buchegger and J.-Y. Le Boudec, “Performance Analysis of the CONFIDANT Protocol: Cooperation of Nodes, Fairness in Dynamic Ad-Hoc Networks,” Proc. MobiHoc, June 2002.
[13] Sanjeev Rana, Manpreet Singh, “Performance Analysis of Malicious Node Aware Routing for MANET using Two-Hop Authentication”, International Journal of Computer Applications (0975 – 8887), Volume 25– No.3, July 2011
[14] Rakesh Kumar, Piush Verma, Yaduvir Singh, “Design and Development of a Secured Routing Scheme for Mobile Adhoc Network”, International Journal of Computer Applications (0975 – 8887) Volume 13– No.2, January 2011
[15] Rajib Das, Bipul Syam Purkayastha, Prodipto Das, “Security Measures for Black Hole Attack in MANET: An Approach”, International Journal of Engineering Science and Technology (IJEST), 2011

[16] S. Kannan, T. Maragatham, S.Karthik, V.P. Arunachalam, "A study of Attacks, Attack Detection, and Prevention Methods in Proactive and Reactive Routing Protocols", International Business Management, 2011
[17] Aishwarya Sagar, Anand Ukey, Meenu Chawla, "Detection of Packet Dropping Attack Using Improved Acknowledgement Based Scheme in MANET", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010
[18] Sowmiya Hariharan, Jothi Precia, Suriyakala.C.D, Prayla Shyry, "A Novel Approach for Detection of Routes with Misbehaving Nodes in Manets", International J. of Recent Trends in Engineering and Technology, Vol. 3, No. 2, May 2010
[19] Usman Yaseen, Ali Zahir, Faraz Ahsan, and Sajjad Mohsin, "Estimating the Effects of Jammers via Conservation of Flow in Wireless AdHoc Networks", International Journal for Advances in Computer Science, Volume 1, Issue 1, 2010
[20] Sohail Abbas, Madjid Merabti, and David Llewellyn-Jones, "A Survey of Reputation Based Schemes for MANET", The 11th Annual Conference on The Convergence of Telecommunications, Networking & Broadcasting (PGNet 2010), Liverpool, UK, 21-22 June 2010
[21] Pradip M. Jawandhiya et. al. / International Journal of Engineering Science and Technology Vol. 2(9), 2010, 4063-4071.
[22] Nishu Garg and R.P.Mahapatra, "MANET Security Issues ," IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.8, August 2009.
[23] B. Kannhavong, H. Nakayama, N. Kato, A. Jamalipour, and Y. Nemoto. A study of a routing attack in OLSR-based mobile ad hoc networks. International Journal of Communication Systems, 20(11):1245–1261, 2007.
[24] B. Kannhavong, H. Nakayama, N. Kato, Y. Nemoto, and A. Jamalipour. A collusion attack against olsr-based mobile ad hoc networks. In GLOBECOM, 2006.
[25] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato, and A. Jamalipour. A Survey of routing attacks in mobile ad hoc networks. IEEE Wireless Communications, page 86, 2007.
[26] Y. Hu, A. Perrig, and D. Johnson, "Ariadne: A Secure On-Demand Routing for Ad Hoc Networks," Proc. of MobiCom 2002, Atlanta, 2002.
[27] Y. Hu, D. Johnson, and A. Perrig. SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks. Ad Hoc Networks, 1(1):175–192, 2003.
[28] Y. Hu, A. Perrig, and D. Johnson. Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks. Wireless Networks, 11(1):21–38, 2005.
[29] C. Tseng, S. Wang, C. Ko, and K. Levitt. Demem: Distributed evidence driven message exchange intrusion detection model for manet. In Recent Advances in Intrusion Detection, pages 249–271. Springer, 2006.
[30] C. Tseng, T. Song, P. Balasubramanyam, C. Ko, and K. Levitt. A Specification-Based Intrusion Detection Model for OLSR. LECTURE NOTES IN COMPUTER SCIENCE, 3858:330, 2006.
[31] M.Wang, L. Lamont, P. Mason, and M. Gorlatova. An effective intrusion detection approach for OLSR MANET protocol. In Secure Network Protocols, 2005.(NPSec). 1st IEEE ICNP Workshop on, pages 55–60, 2005.
[32] T. View. Information theoretic framework of trust modeling and evaluation for ad hoc networks. Selected Areas in Communications, IEEE Journal on, 24(2):305–317, 2006.
[33] M.K.Jeya Kumar, R.S.Rajesh, Performance Analysis of MANET Routing Protocols in Different Mobility Models, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.2, February 2009
[34] Abdul Hadi Abd Rahman, Zuriati Ahmad Zukarnain, Performance Comparison of AODV, DSDV and I-DSDV Routing Protocols in Mobile Ad Hoc Networks, European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4 (2009), pp.566-576
[35] Dipankar Deb, Srijita Barman Roy, and Nabendu Chaki, LACBER: A new location aided routing protocol for GPS scarce MANET, International Journal of Wireless & Mobile Networks (IJWMN), Vol 1, No 1, August 2009
[36] E.A.Mary Anita, V.Vasudevan, Black Hole Attack on Multicast Routing Protocols, Journal of Convergence Information Technology Volume 4, Number 2, June 2009
[37] Ashwani Kush, P. Gupta and C.Jinshong. Hwang, Secured Routing Scheme for Adhoc Networks, International Journal of Computer Theory and Engineering, Vol. 1, No. 3, August, 2009 1793-8201
[38] Sheenu Sharma, Roopam Gupta, Simulation study of blackhole attack in the mobile ad hoc networks, Journal of Engineering Science and Technology Vol. 4, No. 2 (2009) 243 – 250
[39] Cong Hoan Vu, Adeyinka Soneye, An Analysis of Collaborative Attacks on Mobile Ad hoc Networks, Master Thesis Computer Science Thesis no: MCS-2009:4 June 2009
[40] Irshad Ullah, Shoaib Ur Rehman, Analysis of Black Hole Attack on MANETs Using Different MANET Routing Protocols, Master Thesis Electrical Engineering Thesis no: MEE 10:62 June, 2010
[41] Shishir K. Shandilya, Sunita Sahu, A Trust Based Security Scheme for RREQ Flooding Attack in MANET, International Journal of Computer Applications (0975 – 8887) Volume 5– No.12, August 2010
[42] Akanksha Saini, Harish Kumar, Effect Of Black Hole Attack On AODV Routing Protocol In MANET, IJCST Vol. 1, Iss ue 2, December 2010
[43] Aishwarya Sagar, Anand Ukey, Meenu Chawla, Detection of Packet Dropping Attack Using Improved Acknowledgement Based Scheme in MANET, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010
[44] Moitreyee Dasgupta, S. Choudhury, N. Chaki, Routing Misbehavior in Ad Hoc Network, 2010 International

Journal of Computer Applications (0975 - 8887) Volume 1 – No. 18

[45] S. Kannan, T. Maragatham, Attack Detection and prevention methods in Proactive and Reactive Routing protocols, International Business Management 5(3), 2011

[46] Amrit Suman, Praneet Saurabh, Bhupendra Verma, A Behavioral Study of Wormhole Attack in Routing for MANET, International Journal of Computer Applications (0975 – 8887) Volume 26– No.10, July 2011

[47] H. Yang, H. Luo, F. Ye, S. Lu, L. Zhang, "Security in mobile ad hoc networks: challenges and solutions," In proc. IEE Wireless Communication, UCLA, Los Angeles, CA, USA; volume- 11, Page(s):38- 47, ISSN: 1536-1284

[48] Ping Yi, Yue Wu and Futai Zou and Ning Liu, "A Survey on Security in Wireless Mesh Networks", Proceedings of IETE Technical Review, Vol. 27, Issue 1, Jan-Feb 2010.

[49] Y. Hu, A. Perrig, and D. Johnson, "Packet Leashes: A Defense Against Wormhole Attacks in Wireless Ad Hoc Networks," Proc. of IEEE INFORCOM, 2002

[50] K. Sanzgiri, B. Dahill, B.N. Levine, C. Shields, E.M. Belding-Royer, "Secure routing protocol for ad hoc networks," In Proc. of 10th IEEE International Conference on Network Protocols, Dept. of Comput. Sci., California Univ., Santa Barbara, CA, USA. 12-15 Nov. 2002, Page(s): 78- 87, ISSN: 1092-1648.

[51] H. Deng, W. Li, Agrawal, D.P., "Routing security in wireless ad hoc networks," Cincinnati Univ., OH, USA; IEEE Communications Magazine, Oct. 2002, Volume: 40, page(s): 70- 75, ISSN: 0163-6804

[52] D.Suresh kumar, K.Manikandan, M.A.Saleem Durai, Secure On-Demand Routing Protocol for MANET using Genetic Algorithm, International Journal of Computer Applications (0975 – 8887) Volume 19– No.8, April 2011

[53] Zahra Moradi, Mohammad Teshnehlab, Intrusion Detection Model in MANETs using ANNs and ANFIS, 2011 International Conference on Telecommunication Technology and Applications Proc .of CSIT vol.5 (2011) © (2011) IACSIT Press, Singapore

[54] James Cannady, Dynamic Neural Networks In The Detection Of Distributed Attacks In Mobile Ad-Hoc Networks, International Journal of Network Security & Its Application (IJNSA), Vol.2, No.1, January 2010

[55] Nauman Mazhar, Energy Efficient Security in MANET: A comparison of Cryptographic and Artificial Immune System, Pak. J. Engg. & Appl. Sci. Vol. 7, Jul., 2010 (p. 71-94)

[56] Maha Abdelhaq, Rosilah Hassan, Mahamod Ismail, Raed Alsaqour, Daud Israf, Detecting Sleep Deprivation Attack over MANET Using a Danger Theory –Based Algorithm, International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(3): 534-541 The Society of Digital Information and Wireless Communications, 2011 (ISSN: 2220-9085)

[57] Aikaterini Mitrokotsa, Manolis Tsagkaris and Christos Douligeris, Intrusion Detection in Mobile Ad Hoc Networks Using Classification Algorithms, Intrusion Detection in Mobile Ad Hoc Networks Using Classification Algorithms. CoRR, 2008.

# Performance Tuning of Data Transfer in Vehicular Networks

**Dinakaran M[1], Dr. P. Balasubramanie [2]**
**[1] School of Information Technology & Engineering, VIT University,**
**Vellore, Tamil Nadu, 632014, India**

**[2] Department of CSE, Kongu Engineering College,**
**Perundurai, Tamil Nadu, 638015, India**

**Abstract**
Uninterrupted connectivity to Internet services during mobility is an increased demand nowadays. When a user is traveling through a vehicle like car or train, it is important to have a continuous Internet connection without connection break-ups and data losses. This vehicular Internet connectivity is enabled by establishing connectivity to nearby netwoks.N-PMIPV6 Novell Architecture enables seamless and efficient integration of mobile networks. Users moving through vehicles access Internet through the mobile networks. In case the user access multimedia data such as audio, video, graphics and animation, this type of data requires high bandwidth in mobile networks. Accessing such high proportion of data through wireless networks may degrade the quality of data through data loss. The quality of audio or resolution of the downloaded video may get affected due to significant data loss in mobile wireless networks. This article proposes a solution to avoid the data loss issue by simultaneous binding. The proposal is implemented using NS2 simulation tool and the results are analyzed.

***Keywords:*** *NEMO, PMIPv6, Simultaneous Binding, Vehicular Networks.*

## 1. Introduction

Uninterrupted continual flawless Internet service is the demand of the user in today's environment. The most required Internet service is the download or transfer of multimedia data. It is obvious that any data or packet loss that occurs during data transmission affects the quality of the multimedia data. Today the usage of internet is increasing and quality of data transmission is the critical issue under consideration. Many techniques are currently being implemented to provide Internet through vehicular networks however lossless data transfer is the key factor to be considered. Mobile IP is the Internet Protocol for wireless mobile networks developed by Internet Engineering Task Force [1]. When a node changes its point of attachment, this protocol informs the network about the details regarding the switching of host to a new access network. It provides location-independent access to Internet. Each mobile host is assigned a home address when it is present inside the home network. When the mobile host moves out of the home network, it is identified by the Care of Address (CoA) which is registered with the home agent (HA) [2]. Mobile IP specifies the procedure of how a mobile node registers its CoA with the home agent and how the home agent routes the packets to the mobile node through an authenticated procedure [3]. Mobile IP is mostly used in wireless networks where mobile devices traverses across multiple LANs.

## 2. Network Mobility (NEMO)

To support the movement of a complete network which changes its point of attachment to the fixed infrastructure and to maintain the sessions of the mobile nodes uninterrupted, the concept of Network Mobility (NEMO) was developed by the IETF [12]. This provides the mobility management at the network layer [4]. When the mobile node moves out of the home network, it sends a request for a care of address from the new access router. After receiving the new CoA, mobile router has to register this CoA to the home agent. After the successful registration of new CoA with the home agent, a bi-directional tunnel is established between the HA and the mobile router (MR) with one end point as HA's Home Agent's address and the other end point being the MR's Care of Address. This concept is called as tunneling and all the messages to the mobile node are sent through this tunnel. In case of network in a network, a phenomenon called nested tunneling is established. In nested tunneling though multiple tunnels are established in between the mobile node and home agent based on level of nesting, the ultimate tunnel end points are the home agent and the present CoA of mobile node [5]. The advantage of using NEMO is less power consumption for connecting to a particular device outside the vehicle. The mobility of many users can be tracked using a single gateway, as it provides mobility management for them [13].

## 3. Proxy Mobile IPv6

PMIPV6 is a protocol devised by IETF to support mobility for a large network [6]. This protocol makes use of Mobile Access Gateway (MAG) [14] whose functionality is similar to that of access routers. MAG helps in providing connectivity to the mobile terminals present within its range. This protocol makes use of Localized Mobility Anchor (LMA) [7]. LMA acts as a proxy and takes care of several mobile access gateways within its coverage area. The following Figure – 1 illustrates PMIPv6 concept.

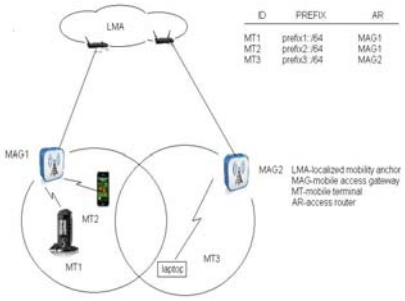



Figure 1 – PMIPv6

When a mobile terminal reaches a new mobile network, it requests a new CoA to the MAG within its range. MAG allocates the corresponding prefix to the mobile terminal and the information regarding the new mobile terminal along with its allocated prefix is registered and stored in LMA in the form of table. LMA acts as proxy and takes up the responsibility of informing the new CoA to the mobile terminal's Home Agent (HA). After CoA binding, a tunnel is established between the LMA and HA. If the mobile terminal starts moving and reaches a new MAG which is under the control of same LMA as the previous MAG, new CoA is updated only in LMA table and details regarding this localized mobility are not informed to the home agent. This is because MAG acts as proxy and takes the responsibility of tracking and updating the location of the mobile node to the central authority. However when the mobile terminal moves to a MAG under the control of a different LMA, this movement has to be informed to the home agent and registered. This protocol acts as an existing mechanism to control mobility of terminals in a large network [8].

## 4. Simultaneous Binding

Simultaneous Bindings is an extension of FMIPv6 [9]. This particular technique minimizes the packet loss to MN. The traffic of the MN is sent to its current location and the next expected location [10]. In vehicular networks, the terminals are in continual motion and hence it is complex to track the location of the terminals at specified point of time. Under many circumstances the mobile terminal moves to a new location and receives a new CoA while the packets are still delivered to the previous access routers. This problem is due to the binding delay of the new CoA to the home agent.

A simple solution for the above problem is to bi-cast or n-cast the data packets for a short period of time from OAR (old access router) to one or more future locations before the node reaches it.

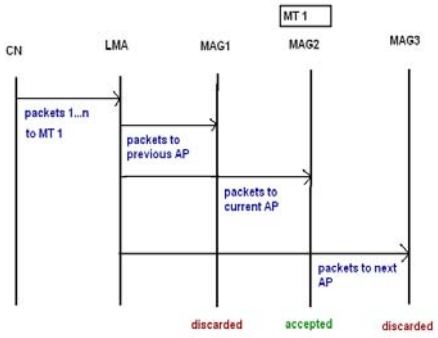



Figure 2 – Simultaneous binding flow

Figure 2, the sequence diagram illustrates the concept of simultaneous binding using three gateways and one mobile node. The above sequence diagram (Figure 2) shows the concept of simultaneous binding using three gateways and one mobile node. Consider a scenario where the data transmission takes place between the mobile terminal and corresponding node. Let us assume that the mobile terminal moves out of the coverage area of MAG1 and reaches the range of MAG2. In normal conditions, the packets are still transmitted to the old MAG without the knowledge of change in location. However in case of simultaneous binding, the packets are transmitted to the present and nearby future locations with a short life time [9]. According to this concept, the packet is first transmitted to MAG1, MAG2 and MAG3. Only the current AP accepts the packets while the rest AP's discard the packets at the expiry of the life time. Unlike other techniques which demands more information to be processed at each component like storing the address, location and other details of the user, this technique requires the storage of less bandwidth table information. When the network tries to reduce the latency, fast handover and HMIPv6 are good [15]. To avoid packet loss, simultaneous binding is a better solution [11].

## 5. NPMIPv6 (NEMO enabled PMIPv6)

In this approach, network mobility is integrated with the usage of MAG and LMA [8]. Using this approach, users can obtain connectivity either from fixed locations or mobile platforms (e.g., vehicles) and obtain continual data transmission irrespective of the change in location. N-PMIPv6 architecture exhibits two remarkable characteristics. First, N-PMIPv6 is totally network-based therefore no mobility support is required in the terminals. This means that the mobility of terminals within the network is transparent. Second, the handover performance is improved, both in terms of latency and signaling overhead.

NEMO Basic Support protocol requires MR (mobile router) to manage their mobility; this is not required in N-PMIPv6. N-PMIPv6 makes use of LMA and MAG to manage the mobility of entire network and hence mobile routers to manage the individual mobile terminals are not required. The LMA-localized mobility anchor adds the new binding cache entry associating the id of MT with prefix. LMA table also contains details regarding the MAG to which the MT is being attached. The MAG acts as a proxy for the mobile node and so only one control message is sent to the LMA. The disadvantage of this approach is the nested tunneling that in turn leads to packet loss. This packet loss degrades the quality of multimedia data that is transmitted.

Though this architecture provides connectivity for the vehicular networks, there is a security issue regarding the authentication of CoA given by MAG to LMA table. Providing the optimized route to the vehicular networks is another big challenge for this existing architecture. The major drawbacks are,

1. Nested tunnels while delivering packets to the mobile nodes.
2. The major concern is about packet loss which degrades the performance especially in terms of multimedia data transmission.
3. Mobile access gateways should be authenticated by LMA's.

## 6. Proposed Architecture

The proposed architecture integrates N-PMIPV6 protocol with simultaneous binding. N-PMIPv6 uses MAG and LMA to provide continuous connectivity while simultaneous binding eliminates packet loss during mobility of terminals by n-casting the data packets to many access routers.

The proposed architecture, which is in Figure - 3 can be explained as follows. Let us consider an ongoing data transmission between a CN and a mobile terminal. Data packets from CN are first transmitted to the home agent of the mobile terminal. The packets are then transmitted from HA to corresponding LMA through the bi-directional tunnel. The concept of simultaneous binding is integrated during the transmission of packets from LMA to MT [16]. For this purpose, the nodes that are under constant motion are identified. Simultaneous binding concept is applied to these terminals under constant motion. LMA transmits the packets to current MAG and future MAG's within its coverage. These data packets are provided with a short life time. This approach makes sure that packet loss to MT is prevented. This is because even if MT reaches a nearby MAG, packets are delivered since the packets are n-casted by LMA to present and nearby MAGs. Only the MAG which contains the MT within its range accepts the packets while the remaining MAGs discard the packets. Duplication of the packet is being avoided by the usage of life time. Each and every packet is sent along with its life time. In case the packet is not delivered to the destined node within the given life time, the packet gets discarded. Thus packet loss as well as packet duplication is being eliminated in this approach. The problem of packet loss in N-PMIPV6 is eliminated that makes reception of good quality video and audio files.

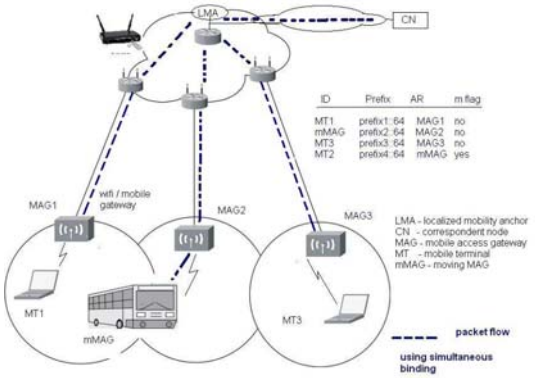



Figure 3 – Proposed Architecture

This is the main table (Table 1) maintained at LMA and acts as a root for all the data transfer. The parameters in the table are id of the particular node, address of the node ipv6, access router which it comes under and the m flag for indication that the user is constantly moving using the vehicle. This information inside the table is maintained at the LMA-local mobility anchor and the other node constantly updates it with the necessary information whenever a little change is made to the network. The information is shared by the trusted node or MAG- mobile access gateway. If the CN needs to send data to the particular node which comes under the LMA then it verifies the table and sends the packet exactly to the destination node using the MAG. MAG plays a vital role

in filling up the above table. If a mobile node comes under the particular MAG, it sends its IPV6 address to the LMA without the usage of bandwidth from the node. So that the proxy updating concept is implemented and the node in the network gets the more connectivity.

| ID | PREFIX | AR | M-FLAG |
|---|---|---|---|
| MT1 | Pref1::/64 | MAG1 | No |
| mMAG1 | Pref2::/64 | MAG1 | No |
| MT2 | Pref3::/64 | MAG3 | No |
| MT3 | Pref4::/64 | mMAG1 | Yes |

Table 1 – LMA table

If the node is constantly moving in the vehicle at some speed in such cases the m-flag is set as true and rest for all the cases the m-flag is false. Using m-flag the mobility of the node can be tracked and also the data transfer is much simpler than before.

| PACKET ID | LIFE TIME (S) |
|---|---|
| P1 | 34S |
| P2 | 20S |
| P3 | 10S |

Table 2 - MN Table

The Table-2 is maintained at the mobile node itself so that it can avoid the duplicate packets. When the mobile node is between the two MAG's it may get many duplicate packets and it makes the node congested. The parameters are packet id and lifetime, using lifetime alone the mobile node rejects the duplicate packets.

### 6.1 Advantages

The proposed system enhances the performance of internet access in the vehicular network.

1. The system guarantees continuous data transmission without packet loss which is eliminated using the concept of simultaneous binding.
2. Packet duplication is avoided by the maintenance of a separate MN table consisting of the packet ID and its life time.
3. Handoff latency is greatly reduced by the usage of N-PMIPv6 which makes use of LMA that acts as localized home agent and performs packet reception on behalf of the mobile terminal.
4. The proposed architecture consumes only lesser bandwidth and hence the traffic is greatly reduced.
5. The time taken for reverse tunneling in the proposed architecture is very low when compared to other existing systems.

## 7. Performance Analysis

The Packet loss, bandwidth, traffic/congestion, latency and reverse tunneling are the parameters that are used for the performance analysis. After every simulation, trace file (with .tr extension) is generated by the simulator. On comparing the existing system trace files with the proposed system trace file it is evident that the proposed system is far better than the existing system.

### 7.1 Packet Loss

The number of packets dropped is counted from the trace files of existing system and proposed system. The dropped packets are mentioned by –d in the trace file. Those lines are alienated and counted. A graph is plotted with "simulation time" in X-axis and "number of packet loss" in Y-axis. The overall simulation of existing system and the proposed system are same so as to make the comparison easier. The packets dropped at particular time interval is marked and connected. The existing system has more packet loss whereas the proposed system has fewer packet losses. The values from the trace file are plotted as points in the graph (Figure – 4).

Both the systems are validated with their number of highest packet losses. The existing system shows highest packet loss of 56 packets at time 18, the proposed system shows highest packet loss of 33 packets at time 34 (at worst case scenario-occurs rarely). Thus, the integrity and efficiency of new system is proved. The comparison graph is shown below,

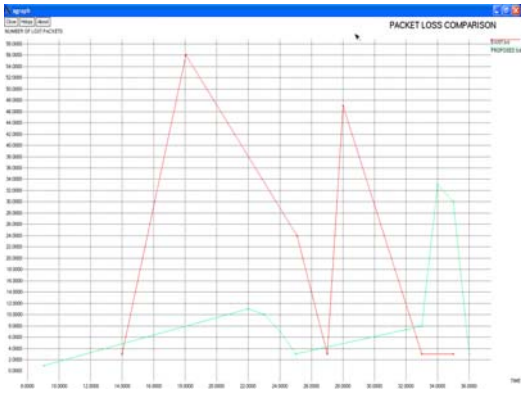



Figure 4 - Comparison graph for packet loss

The red color line in the graph represents the existing system/architecture and the green color line represents the proposed system/architecture.

### 7.2 Bandwidth

The bandwidth is the overall capacity/channels in the particular network. In this case the bandwidth utilization for single data transfer through the network is considered

for plotting the graph. During data transfer some packets are dropped. Those packets are resend by the source on receiving negative acknowledgement. These additional packets are also taken into consideration. Here the graph is plotted with "time" in X-axis and "packets" in Y-axis. Both existing and proposed architecture are plotted in the same graph. From the trace file all categories of packets such as sent and received packets are counted using the word count command.

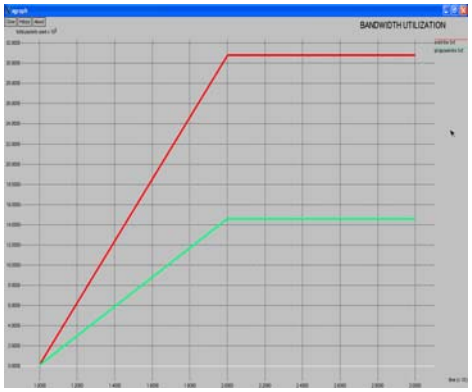



Figure 5 - Bandwidth utilization comparison graph

As mentioned in the (Figure – 5) graph, red line is used to indicate the existing architecture and the green line is used to indicate the proposed architecture. In the proposed system bandwidth consumption is less which means the remaining bandwidth can be used for other purposes. Due to lower bandwidth consumption, the traffic is also greatly reduced.

## 7.3 Congestion/traffic at each node

Congestion of packets denotes the overcrowding of the packets at the nodes. In both the existing and proposed architecture, the same numbers of packets are transferred between two nodes. The number of packets handled at each node is taken for plotting the graph. The "congestion" at each node is taken in X-axis and "packets" is taken in Y-axis. The X-axis is the number of nodes involved in the simulation (1 to 5). Even in congestion graph (Figure – 6) the proposed system proved to be efficient.

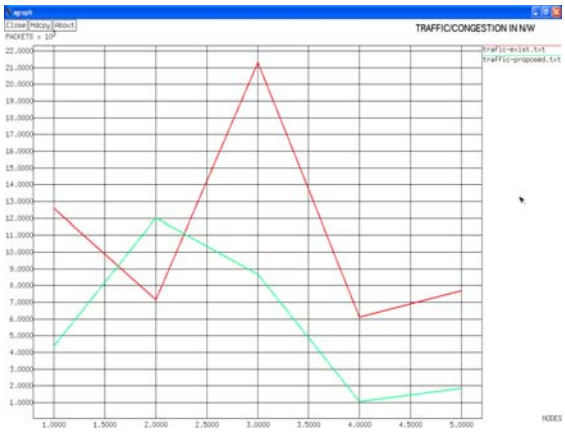



Figure 6 - Traffic/congestion in the network

Hence the proposed system can handle more number of users in its network than the old one and also can manage all the data transfers between them efficiently.

## 7.4 Reverse Tunneling

In vehicular network, the user moves frequently. When some node wants to send packets to user who is constantly moving, the home network sends the packets to the access point which is nearer to the user. When the packets are being transferred, the user moves from one access point to another access point. At this situation, the packets from the previous access point are redirected back to the home network. The home network redirects the packets to the new access point, which is nearer to the user. This is called reverse tunneling. Reverse tunneling occupies more bandwidth and causes more packets lose and so on. In the new system, the occurrence of reverse tunneling is very less or no reverses tunneling. So there will be little packet loss and uses less bandwidth, hence it performs well in terms of data transfer. The "access points" (0, 1, 2 & 3) are taken in X-axis and the "time taken" by the nodes for re-tunneling the packets is taken in Y-axis. The time is considered in seconds. The graph below (Figure – 7) shows the difference between the two systems. For example, user moves at some certain high speed inside the vehicle, it has to travel across many access points. If more packet loss happens, the quality of data will be reduced. Nowadays multimedia data are having high demand and used frequently. For exchange of multimedia data more precision in data exchange is needed. Packet losses must be negligible so that it doesn't affect the sensitivity of the data. Providing a qualitative data exchange is possible with the new system. Reverse tunneling is a rare event in the proposed system, assuring that the multimedia data is transferred with higher integrity.

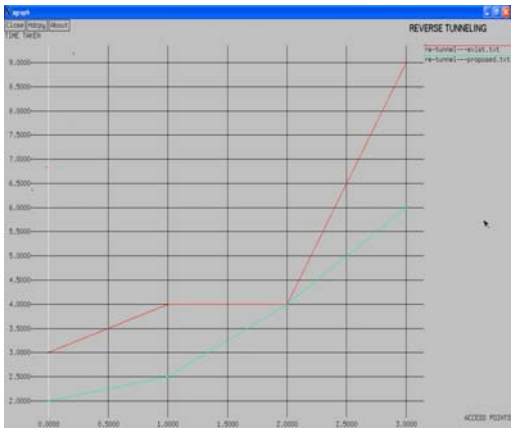


Figure 7 - Reverse Tunneling graph comparison

Now, the same parameter is shown in bar graph (Figure – 8) below along with the latency comparison. Latency is the time taken for data transfer from the source to destination till all packets are reached. Reverse tunneling will increase the latency which reduces the performance so the latency should be very low.

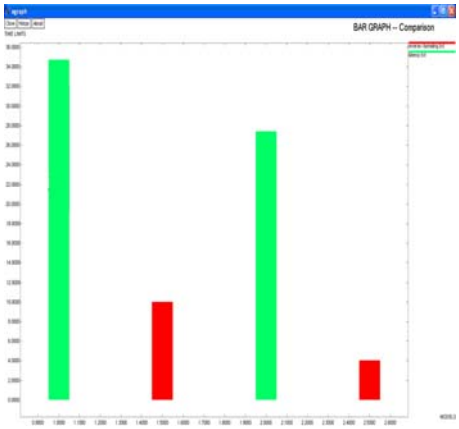


Figure 8 - Bar graph for latency and reverse tunneling

The green color bar represents the latency. The first green color long bar indicates the existing system. The existing system takes more time to reach the destination and the second small green color bar indicates the less latency period in proposed system. Hence the packets from the new system reach the destination faster than the existing system.

The first red color bar represents the time taken for the reverse tunneling in the existing system and then the second red color bar indicates the time taken for reverse tunneling in the proposed system.

## 8. Conclusion

The proposed method improves the performance of the vehicular networks in terms of data transfer. The packet loss is minimized, thus improving the quality of data transfer especially transferring multimedia data which is frequently used by the users. Our architecture will provide route optimization for the vehicular networks. In future, the security problems like authenticating the access points in order to prevent falsifying the table information can be addressed. This new architecture may be implemented in other type of networks.

## References

[1] “Mobile IP”, http://en.wikipedia.org/wiki/MIPv6
[2] Fyza Nada “Performance Analysis of MIPV6 and MIPV4”, the international Arab journal of information technology, Vol. 4 2 , April 2007
[3] X.Pérez-Costa and H.Hartenstein “A Simulation Study on the Performance of Mobile IPv6 in a WLAN-Based Cellular Network” Computer Networks Journal (CNJ), September 2002.
[4] J. Mangues, A. Cabellos, R. Serral, J. Domingo, Gómez, T. de Miguel, M. Bagnulo. A. García. “IP Mobility: Macromobility, Micromobility, Quality of Service and Security,” Journal of the Council of European Professional Informatics Societies, Vol. 5 (1), pp. 49- 55, February 2004
[5] Ben McCarthy, Matthew Jakeman, Dr Chris Edwards, Pascal Thubert , “Protocols to Efficiently Support Nested NEMO (NEMO+)”, MobiArch, August 2008, pp. 43-48
[6] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, “Proxy Mobile IPv6”, Internet Draft, IETF, May 30, 2008
[7] Chin-Chen Chang, Chia Yin Lee b, Yen-Chang Chiu b “Enhanced authentication scheme with anonymity for roaming service in global mobility networks”, Elsevier 2008.
[8] Ignacio Soto, Carlos J. Bernardos, Maria Calderon, and Albert Banchs, “NEMO-Enabled Localized Mobility Support for Internet Access in Automotive Scenarios”, Topics in automotive scenario 2009.
[9] C. Perkins, "IP Mobility Support for IPv4", RFC3220, IETF, Jan 2002
[10] Patrick P. Lam and Soung C. Liew, “Nested Network Mobility on the Multi hop Cellular Network”, IEEE communications magazine 2007.
[11] Karim El Malki and Hesham Soliman, “Simultaneous Bindings for Mobile IPv6 Fast Handovers”, Internet Draft, IETF, July 2005
[12] Neeraj Sharma1, Naveen Sharma2, and Dimple Malik, "Problems of Inefficient Routes in Network Mobility in Wireless Multihop Gateway Networks and

their Resolution", International Journal of Electronics Engineering, 3 (1), 2011, pp. 29– 31
[13] S. Gundavelli, G. Keeni, K. Koide, K. Nagami, "Network Mobility (NEMO) Management Information Base", RFC5488, April 2009
[14] Ignacio Soto, Carlos J. Bernardos, María Calderón, and Telemaco Melia,"PMIPv6: A Network-Based Localized Mobility Management Solution", The Internet Protocol Journal, Volume 13, No.3
[15] Xavier Perez Costa, Ralf Schmitz, Hannes Hartenstein, Marco Leibsch, "A MIPv6, FMIPv6 and HMIPv6 handover latency study : Analytical Approach", June 2002
[16] Dinakaran M, Balasubramanie P,"Integrating N-PMIPv6 and Simultaneous Bindings Avoid Packet Loss in NEMO", International Journal of Computer Applications, vol. 15, issue 4, pp. 33-36, Feb 2011

**M. Dinakaran** has completed his B.Tech (IT) and M.Tech (IT-Networking) in Vellore Institute of Technology, Tamil Nadu and India. He had worked in TATA Consultancy Services as a Assistant System Manager between Sep 2006 to July 2009. He has been awarded as TCS Gems during third quarter of 2008. Currently he is working as Assistant Professor in VIT University, Vellore and he is pursuing Ph. D in Kongu Engineering College, under guidance of Dr. P. Balasubramanie. He has published 10 articles in International Conferences / Journals

**Dr. P. Balasubramanie** has been awarded Junior Research Fellowship (JRF) by CSIR in the year 1990. He completed his PhD degree in 1996 at Anna University, Chennai. Currently he is a professor in the Department of Computer Science and Engineering in Kongu Engineering College, Perundurai, and Tamilnadu, India. He has guided 7 PhD scholars and guiding 20 scholars Under Anna University. He has published more than 70 articles in International/ National Journals/Conferences. He has authored 6 books with the reputed publishers.

# Design of Miniature Patch Antenna Around the Frequency 3.5 GHz for WIMAX Technology

**Adnane-Latif** [1]

[1] **Laboratory of Information Technology and Modeling, Cadi Ayyad University, National School of Applied Sciences, Marrakesh, Morocco**

**Abstract**

This work aims to study a miniature rectangular patch antenna λ / 8 fed by coaxial probe with the transmission line method (TLM). The design and simulation of this antenna is around the frequency of 3.5GHz, for WIMAX technology. The results obtained (input impedance, reflection coefficient, VSWR and bandwidth) are given by the program in the software MATLAB.

***Keywords:*** *Patch antenna, Wireless Communications, Wireless Metropolitain Area Network, Wimax, Design.*

## 1. Introduction

In the first part, this paper deals the modeling of the rectangular patch antenna λ / 8 fed by coaxial probe with the TLM model.

The second part of this paper is devoted to the simulation of this antenna in the frequency of 3.5 GHz for WIMAX applications, to determine the parameters necessary for the antenna's design.

The key parameters of these simulations are: the input impedance, the resonance frequency, the reflection coefficient and the standing wave ratio (VSWR), with and without compensation of inductif effect caused by the coaxial probe inductance used to feed the patch antenna λ / 8.

## 2. Modeling of Patch Antenna λ / 8

Either, the patch antenna λ / 8 represented in the below figure:

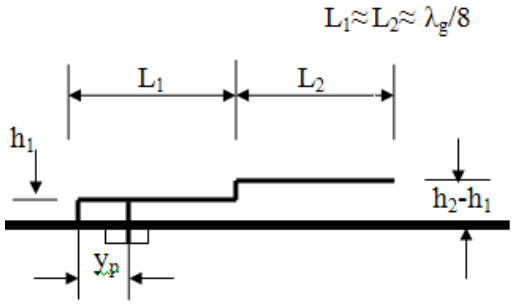



Fig 1 Structure of the patch antenna λ / 8

Its equivalent circuit with the TLM method is shown in the figure 2:

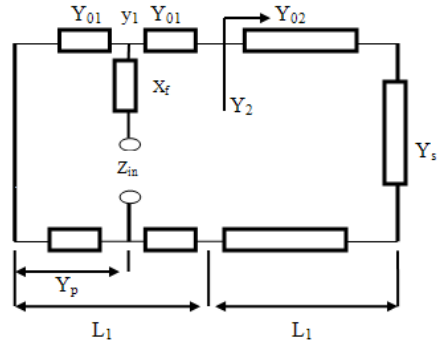



Fig 2 Equivalent circuit of the patch antenna λ / 8 with the TLM method

**$Z_{in}$** : Impédance d'entrée de l'antenne.
**$X_f$** : Réactance inductive
**$Y_{01}$**: Characteristic admittance of the line.
**$Y_{02}$**: Characteristic admittance of the line $L_2$ ($L_1 = L_2$).
**$Z_1$**: Impedance of the antenna without that of the coaxial probe.
**$Z_{in}$**: Input impedance of the antenna.
**$X_f$** : Inductive reactance.

## 3. Design of the Patch Antenna λ / 8

### 3.1 Resonant Frequency and Input Impedance

A design around the frequency of 3.5 GHz for industrial applications in telecommunications namely WMAX, the dimensions of the patch are: The substrate dielectric permittivity $\varepsilon_r$=2.1, the length of the inferior patch is $L_1$ = 14 mm and the upper patch is $L_2$ = 14.5 mm. The widths of the patches W1et W2 are equal and worth 18.7mm. The height of the inferior patch is $h_1$ = 2.78 mm and that of the upper patch $h_2$ = 5.4mm.
The coaxial probe position from the edge of the short circuit is x = 2.2 mm. This gives a resistive part of the input impedance of 49.98 Ω in the resonant frequency 3,5 GHz. The coaxial probe radius is 0.32 mm; the inductance in the resonant frequency is 25.14 nH.

The figure 3 shows the real part of the input impedance, the curve peaks indicates the input resistance of the antenna.

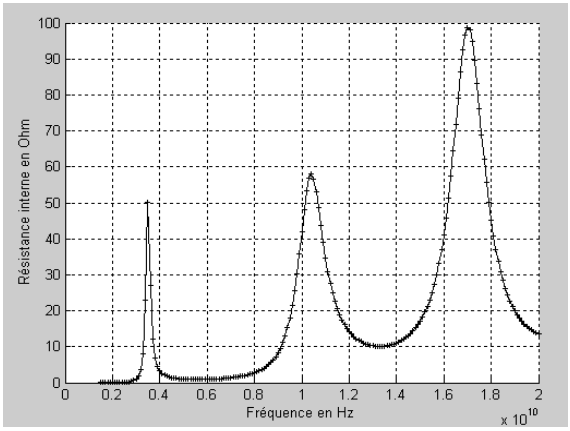



Fig 3 Input resistance of the patch antenna λ / 8 as a function of the frequency.

I can conclude from the figure 3, that resonance frequencies of the patch near 3.5 GHz, 10.4 GHz and 17.2GHz respectively corresponding to the lengths of the patch $L_{1eff}$ = λ / 8, $L_{1eff}$ = 3λ / 8 and $L_{1eff}$=5λ/8 ($L_{1eff}$ is the effective length of the patch) . The fundamental mode of propagation is 3.5 GHz; the other frequencies (10.4GHz and 17.2GHz) are other higher order modes.

The figure 4 below allows us to determine precisely the value of the resonant frequency, that 3.5 GHz which gives the maximum value of the resistive part of the input impedance that is 49.98 Ω at the adaptation point.

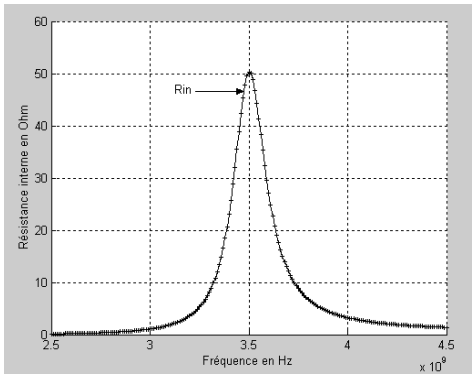



Fig 4 Input resistance of the patch antenna λ / 8 as a function of the frequency that determine precisely the value of the resonant frequency, in the maximum value of the resistive part of the input impedance at the adaptation point

The patch length determines the frequency desired. This length is the sum of the patch length L and the extension of the field lines Δl that gives the resonance frequency.

### 3.2 Impedance and reactance of the antenna input patch.

It can be seen in figure 5, that the input resistance is represents the real part of the input impedance that is 49.98 Ω is (always positive), which corresponds to a resonance frequency of 3.5 GHz.

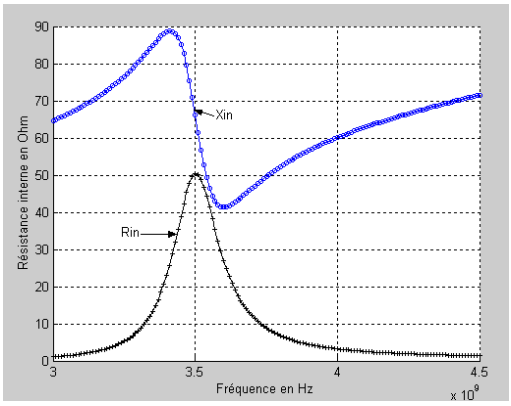



Fig 5 Input resistance of the patch antenna λ/8 as a function of the frequency.

For this frequency there is a imaginary part (reactance) of the input impedance of this antenna that due to the coaxial probe inductance, and represents the metal losses and reflection losses at contact point between the coaxial probe and the patch λ/8.

This reactance has an influence on the adaptation of this antenna. This influence appeared in the reflection coefficient of the antenna as a function of the frequency.

### 3.3 Reflection Coefficient

The following curve shows the reflection coefficient as a function of the frequency:

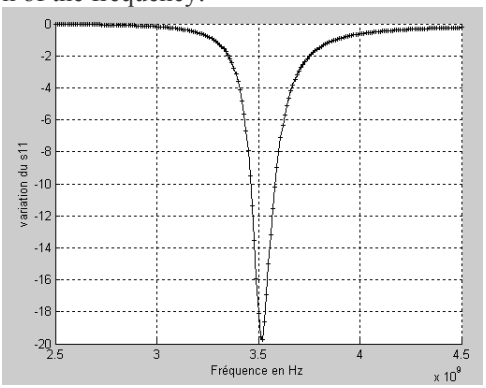



Fig.6: Reflection coefficient as a function of the frequency of the patch antenna λ / 8 before adaptation

Observed on the figure 6, that the reflection coefficient has a value of -19.6 dB at a resonant frequency of 3.52 GHz. It is a low value that shows the mismatch patch antenna caused by the reactance. And the value of the bandwidth of the patch antenna is very small, that is 124 MHz (3.5% of center frequency).

### 3.4 Standing Wave Ratio

The standing wave ratio as a function of frequency is given by the figure 7:

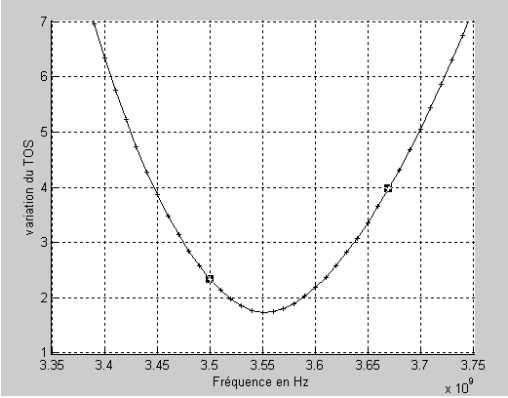



Fig7. Standing wave ratio as a function of frequency of a patch antenna λ / 8 before adaptation.

The SWR has a value of 1.62 at the resonant frequency of 3.55 GHz, is used to calculate the bandwidth for patch antenna but it is low, this due always to the mismatch impedance caused by the reactance.

## 4. Adaptation of the patch antenna λ / 8 with a capacitive reactance.

Let now a capacitor in the middle of the coaxial probe as shown in the figure 8.

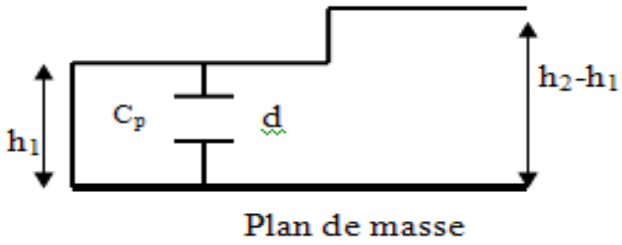



Fig 8 Geometry of the patch antenna with the capacitor.

The equivalent circuit find by the TLM method of this structure is:

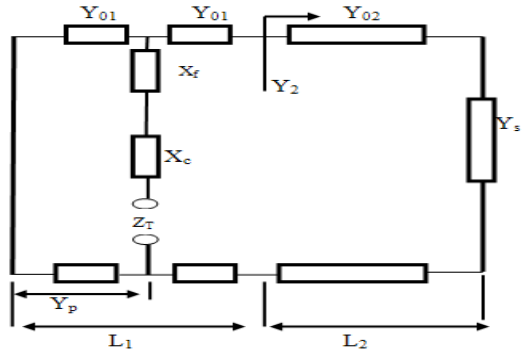



Fig.9 Equivalent circuit of the patch antenna λ / 8 after the addition of capacitor.

The input reactance of this antenna is the sum of the coaxial probe reactance and the reactance that has been added.

We will study the effects of Adding this capacitor on the reflection coefficient and the SWR at the resonant frequency·

### 4.1 Input impedance of the antenna patch

The figure 10 shows the simulation of the patch antenna λ / 8 with the addition of capacitor.

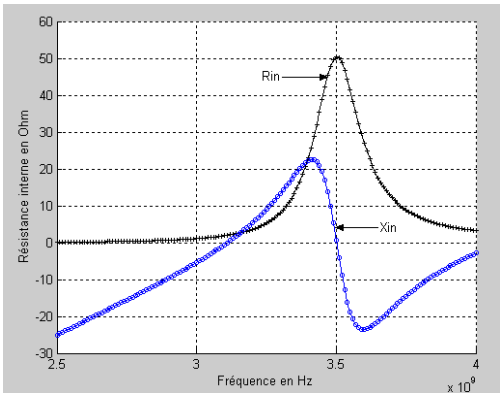



Fig 10 The input resistance and reactance of patch λ / 8 after adaptation.

The figure 10 shows that the input resistance remains the same at the resonant frequency, but the input reactance is almost zero at the resonant frequency, because of the cancellation of the probe coaxial inductive reactance caused by the capacitive reactance.

### 4.2 Reflection coefficient

We observe in the figure 11, an increase of the reflection coefficient with a value of -42.5 dB at the resonant frequency of 3.5 GHz.

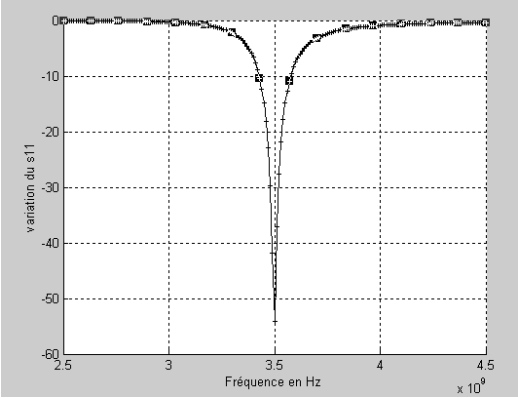



Fig 11 Reflection coefficient as a function the frequency with the addition of capacitor.

Note, that the addition of a capacitor reactance which cancels the effect of inductive reactance causes the decrease of the reflection coefficient of the value -19.6 dB to achieve the value of -54.4 dB at the resonant frequency of 3.5 GHz, this to calculate the bandwidth of the patch antenna.

### 4.3. Standing Wave Ratio (SWR)

The value of the SWR is 1.1 at 3.5 GHz frequency, which shows that our antenna has been adapted (Figure 12).

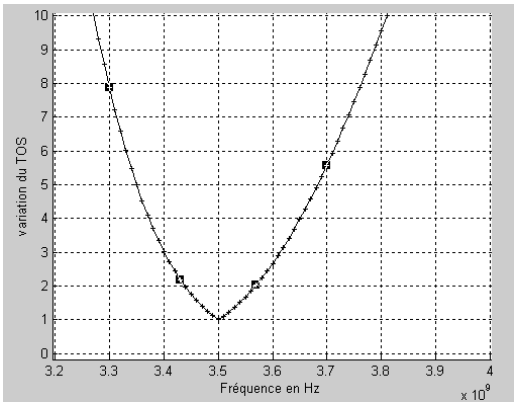



Fig 12 SWR as a function of frequency after the addition of capacitor.

Since the SWR and reflection coefficient at the resonance frequency are respectively less than 2 and -10 dB, then we can calculate the value of the bandwidth of the patch antenna. The value of the bandwidth after the adaptation capacitor is 160 MHz (4.4% of center frequency).

## 5. Conclusion

We conclude that the simulation of the patch antenna λ / 8 at frequency of the WIMAX technology is necessary to improve the reflection coefficient and standing wave ratio. To compensate the reactance due to the inductive effect of the coaxial probe, it can be done by adding capacitor to the dielectric substrate. This increases the bandwidth from 3.5% to 4.4%.

## References

[1] Abdelwaheb Ourir "Applications de matériaux à bandes interdites Photoniques et de matériaux en télécommunications". Thèse de doctorat en sciences : Université de Paris XI d'Orsay, 2006.

[2] J.R. James, and P.S. Hall, " Handbook of microstrip antennas ", I.E.E. Electromagnetic Waves Series 28 - Peter Peregrinus LTD, 1989.

[3]G.TROUILLARD, "contribution à l'étude des phénomènes électromagnétiques lies aux futurs systèmes mobiles de réception hertzienne a bord des véhicules automobiles. Conception, réalisation et tests des antennes correspondantes", Thèse de doctorat en télécommunications : Université de Limoges, France, 2003.

[4] M. DIBLANC, "développement du concept de l'antenne a résonateur BIE pour la génération de la polarisation circulaire ", Thèse de doctorat, université de limoges, France 2006.

[5] N. FAURE-MURET, " Conception, réalisation et tests des filtres millimétriques volumiques micro-usinées'', Thèse de doctorat, Université de Limoges , France, 2005.

[6] L. FRAYTAG, "Conception, réalisation et caractérisation d'antennes pour stations de base des réseaux de télécommunications sans fil", Thèse de doctorat, Université de Limoges. France, 2004.

[7] M.Grzeskowiak, "Antennes multicouches intégrées sur arséniure de gallium à 24 GHz pour applications antennes actives faible portée". Thèse d'Etat, Université de Lille. France, 1999.

**Adnane LATIF** born in El Jadida, Morocco in 1973, Doctor of Telecommunications , and Microwaves from Cadi Ayyad University, is a researcher in the areas of smart antennas, antennas miniaturization, location in mobile networks, W LAN, WWAN. Actually professor in Cadi Ayyad University.

Founder and Coordinator of the Spring School 2010 "Wireless Communications and Emerging Technologies."
Founder and President of the Moroccan Association of Technology, Telecommunications and Electronics
Member organizing partner of the Spring School 2011 "Wireless Communications and Emerging Technologies."
Member Organizer of the International Conference ICMS 2005, 22-24 November 2005, Marrakech, Morocco.
Member of International Program Committee of several international conferences:

-IEEE ICIT 20058-10 December 2004, Tunisia.
-NGN-'10, Marrakesh, United Kingdom, 8-10 July 2010.
-NETAPPS2010) 22-23 September 2010, MALAYSIA.
-NGN-'11, Hammamet, Tunisia, 20-22 May 2011

# Ordinal Classification Method for the Evaluation Of Thai Non-life Insurance Companies

**Phaiboon Jhonpita[1], Sukree Sinthupinyo [2] and Thitivadee Chaiyawat[3]**
[1]Technopreneurship and Innovation Management Program Graduate School, Chulalongkorn University, Bangkok, Thailand

[2]Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand

[3]Department of Statistics, Faculty of Commerce and Accountancy Chulalongkorn University, Bangkok, Thailand

**Abstract**

This paper proposes a use of an ordinal classifier to evaluate the financial solidity of non-life insurance companies as strong, moderate, weak, and insolvency. This study constructed an efficient classification model that can be used by regulators to evaluate the financial solidity and to determine the priority of further examination as an early warning system. The proposed model is beneficial to policy-makers to create guidelines for the solvency regulations and roles of the government in protecting the public against insolvency.

***Keywords:*** *Ordinal classification, Imbalanced class classification, Solvency condition classification, Non-life insurance companies.*

## 1. Introduction

Thailand Insurance industry is subject to government regulation to protect policyholders, third-party liability claimant, and other related business. Solvency supervision, regulations and solvency position classification is an important topic for non-life insurers. Most of the studies were implemented in the United States and many previous studies focused on binary classification and the problem whose class values were unordered (bankrupt/non-bankrupt, solvency/insolvency, or healthy/failed)[2-16]. Unfortunately, they were not implemented in the multi-class classification fashion. In this paper, we hence proposed an ordinal multi-class classification for solvency condition classification. Normally, The Office of Insurance Commission (OIC) of Thailand uses the Capital ratio (CAR) system of non-life insurance in 2009 to evaluate the capital adequacy or financial solidity of the non-life insurers (as shown in Table 1). With the condition distinguished by a level of CAR, the insurance company and regulator's actions are required.

**TABLE 1** The solvency evaluation and regulatory actions based on CAR system.

| Class Classification | Capital adequacy ratio (CAR) | The action level |
|---|---|---|
| Strong | ≥ 150% | No action level |
| Moderate | 120 - 150% | Company action level |
| Weak | 100 - 120% | Regulatory action level |
| Insolvency | < 100% | Authorized control & Mandatory control level |

Note: Company action level - company must file plan with insurance commissioner & explaining cause of deficiency and how it will be corrected. Regulatory action level - The commissioner is required to examine the insurer and take corrective action, if necessary. Authorized control level & Mandatory control level - The commissioner has legal grounds to rehabilitate or liquidate the company, the commissioner is required to seize a company.

The level of capital adequacy ratio (CAR) of insurer is affected by most insurance activities and decision making processes such as premium rate making, determination of the technical reserve, risk undertaking, reinsurance activities, investment, sales, credibility of company to related party, and also be affected by the country's economy, new legislations, inflation and interest rates [1]. With the help of our system, the companies can early detect the solvency condition of their own and can decide the most suitable policy to reduce their risk.

## 2. Literature review

Among many empirical studies of insurance science, there are several studies with different techniques used for improving the performance of Insolvency prediction and/or classification model. Most studies applied traditional statistic techniques, such as regression analysis [2], multivariate discriminant analysis (MDA) [3, 4, 5], logistic regression (LR) [6], logit and probit model [7-10], and multinomial logistic regression (MLR) [1]. On the other hand, machine learning techniques such as neural networks (NNs) [11-15], and genetic algorithm (GA) [16] were also used in Insolvency prediction.

Kramer (1997) evaluated the financial solidity of Dutch non-life insurance by combining a traditional statistic technique (ordered logit model) with artificial intelligence techniques (a neural network and an expert system). The complete model contains three programs; logit model, neural network, and expert system. The data from year 1992 has been used as training data set and year 1993 as the test set. The output of the multi-class classification model consists of the priority for further examination (High, Medium, and Low class). The system which evaluates the financial solidity can be used to classify the insurers according to their degree of risk exposures. The model correctly classified 93% of the data test set. It showed very good performance for strong, medium and weak companies, 96.3% of the strong, 75.0% of the medium and 94.4% of the weak are classified correctly.

Pitselis (2009) studied the solvency supervision, regulations and insolvency prediction of Greece insurance companies using statistical methodologies, e.g. discriminant analysis (DA), logistic regression (LR), and multinomial logistic regression (MLR) to distinguish solvency position into two cases; two-class classification (healthy and insolvency) and multi-class classification (healthy, merged, and insolvency). The paper presented the effects of solvency position of insurance companies. Company and regulatory actions are required if a company's solvency position falls below requirement. Due to the imbalanced data problem, especially for insolvency companies, LR and MLR failed to give reliable results. DA model was able to adequately classify Healthy, Merged, Insolvency companies; 93.5%, 33.3% and 100% respectively (on the 1998 data set).

### 2.1 A Simple Approach to Ordinal Classification

Frank and Hall (2001) [17] presented an ordinal classification approach that enables standard classification algorithms to classify the ordinal class problems. Frank and Hall applied standard classifier in conjunction with a decision tree learner. The underlying learning algorithm takes advantage of ordered class values. First, the original dataset problem is transformed from a $k$-class $V = \{v_1 \ldots v_k\}$ to $k$ -1 binary-class problems. The training starts by deriving new datasets from the original dataset, one for each of the $k$-1 new class attributes. In the next step, the classification algorithm is applied to generate a model for each of the new datasets. To predict the class value of an unseen instance, we need to estimate the probabilities of the $k$ original ordinal classes using our $k$-1 model. Estimation of the probability for the first and last ordinal class value depends on a single classifier**.**

In General, for class values $V_i$, a probabilities distribution on $V_i$ ($k$-classes) is then derived as follows:

$$Pr(V_1) = 1 - Pr(\text{Target} > V1)$$
$$Pr(V_i) = max\{ Pr(\text{Target} > V_{i-1}) - Pr(\text{Target} > V_i), 0 \}, 1 < i < k$$
$$Pr(V_k) = 1 - Pr(\text{Target} > V_{k-1})$$

To classify an instance of an unknown class, the instance is evaluated by each of the $k$-1 classifiers and the probabilities of each the $k$ ordinal class value is calculated using method above. The class with maximum probability is assigned to that instance.

### 2.2 Decision Tree Learning Algorithm

The Decision Tree Learning (DTL) algorithm we used in this research is the one named J48 implemented in WEKA machine learning tool [18]. The J48 class is implemented based on the same concept as C4.5 decision tree [19].

The DTL is a predictive machine learning model which begins with a set of the whole training examples. It creates a decision tree based on the attribute values of the training data that can best classify the set of samples at a time. The attribute which can best discriminate the sample set is evaluated based on the concept of Entropy. The examples are then divided into edges which is the value of the attribute. The child node which consists of examples from different classes will be replaced with the new attribute node, while the child node containing examples from the same class will be a used as a decision node, in which all examples will be classified as the class of training examples collected in this node.

## 3. Data and Methodology

The data set used in this study was collected from 70 non-life insurance companies in Thailand. The companies which were in operation or went insolvency were covered from 2000 to 2008. During this period, 616 cases (543 strong, 16 moderate, 13 weak and 44 insolvency) were selected as training data set as shown in Table 2. The data of year 2009

**TABLE 2** Number of Non-life Insurance companies in this study (Data from year 2009 are the separated test set).

| Class | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | Total | % | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Insolvency | 5 | 3 | 6 | 5 | 6 | 4 | 6 | 5 | 4 | 44 | 7.1% | 6 |
| Weak | 1 | 1 | 1 | 1 | 2 | 0 | 3 | 1 | 3 | 13 | 2.1% | 1 |
| Moderate | 0 | 1 | 1 | 2 | 1 | 4 | 3 | 3 | 1 | 16 | 2.6% | 1 |
| Strong | 64 | 65 | 62 | 62 | 61 | 60 | 56 | 56 | 57 | 543 | 88.1% | 57 |
| Total | 70 | 70 | 70 | 70 | 70 | 68 | 68 | 65 | 65 | 616 | 100 % | 65 |

Note: The solvency condition in this study is determined by capital adequacy ratio = Total capital available (TCA) / Total capital required (TCR)

were used as a separated test set. The data source comes from the annual report of The Office of Insurance Commission (OIC) and the health insurance companies are not including on this study

The attributes selection started from 13 attributes. We chose them from the most commonly used ones in empirical studies of insurance science. They were found significant in previous studies of predicting non-life insurances' solvency [1-11, 13-16]. In this paper, we select the relevant attributes using the correlation-based attribute subset evaluator and greedy stepwise. All 13 attributed are shown in Table 3.

**TABLE 3** Attributes used in this study

V1 Net premiums written / policyholders' surplus
V2 Solvency margin to minimum required solvency margin
V3 Policyholders' surplus & Technical reserve to net written premium
V4 Claims incurred to policyholders' surplus & technical reserve
V5 Gross agent's balance to policyholders' surplus
V6 Change in policyholders' surplus
V7 Investment yield
V8 Investment assets to Policyholders' surplus
V9 Return on total assets (ROA)
V10 Loan & other investment to policyholders' surplus
V11 Loss reserve & unpaid losses to policyholders' surplus
V12 Capitalization ratio
V13 Auto lines net written premium to total net written premium

After we analyzed the distribution of the training data, we found that the distribution of the data set was imbalanced, as shown in Table 2**.** The classification of data with imbalanced class distribution has posed a significant drawback on the performance of most standard classifiers, which assume a relatively balanced class distribution and equal misclassification costs [20]. Many techniques were proposed to solve this problem, for example, re-sampling methods for the balancing the data set, modification of existing learning algorithms, measuring the classifier performance in imbalance domains, relationship between class imbalance, and other data complexity characteristics [21].

To attack the imbalanced data set problem, we employ the standard resample technique to produce a new random set of data by sampling with replacement. The distribution on the data sets after applying resample techniques is presented in Table 4. In this study, we use the ordinal class classifier which employs the DTL algorithm as the base classifier.
Figure 1 shows the classification process. Fig. 2 and 3 shows the concept of testing approaches, 10 fold cross-validations and 70:30% split data set validation.

**TABLE 4** Training data set after applying resample technique.

| Class Classification | Original data set | | Resample data set | |
|---|---|---|---|---|
| Insolvency | 45 | 7.3% | 157 | 25.5% |
| Weak | 13 | 2.1% | 137 | 22.2% |
| Moderate | 17 | 2.8% | 144 | 23.4% |
| Strong | 541 | 87.8% | 178 | 28.9% |
| Total | 616 | 100.0% | 616 | 100.0% |

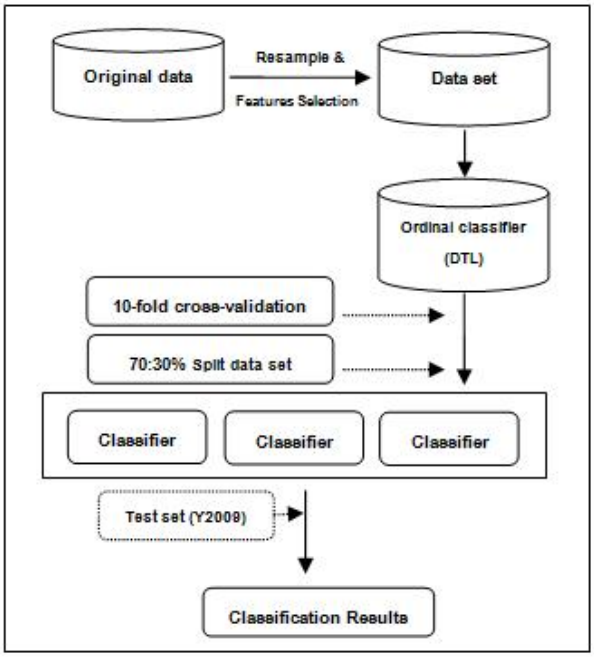



**Fig.1** Model Construction.

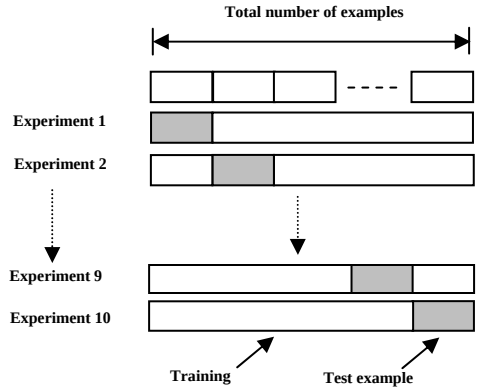



**Fig.2** 10-fold cross-validation

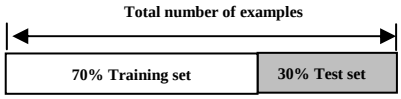



**Fig.3** 70:30% Split data set

## 4. Experimental and Results

This paper used a 10 fold cross-validation, 30% split test set and separated test set (2009 data set). The classification results are shown in Table 5, 6, and 7.

**TABLE 5** Classification results obtained from 10-fold cross-validation (total 616 instances)

| Class Classification | I | W | M | S | Total | Classified Correctly (%) |
|---|---|---|---|---|---|---|
| I | 154 | 3 | 0 | 0 | 157 | 98.1% |
| W | 0 | 137 | 0 | 0 | 137 | 100.0% |
| M | 0 | 0 | 144 | 0 | 144 | 100.0% |
| S | 0 | 0 | 5 | 173 | 178 | 97.2% |
| Total | | | | | 616 | 98.7% |

I = insolvency, W = weak, M= moderate, S= strong

**TABLE 6** Classification results from 30% spilt test set (total 185 instances)

| Class Classification | I | W | M | S | Total | Classified Correctly (%) |
|---|---|---|---|---|---|---|
| I | 49 | 2 | 0 | 0 | 51 | 96.1% |
| W | 0 | 44 | 0 | 0 | 44 | 100.0% |
| M | 0 | 0 | 40 | 3 | 43 | 93.0% |
| S | 0 | 1 | 2 | 44 | 47 | 93.6% |
| Total | | | | | 185 | 95.7% |

I = insolvency, W = weak, M= moderate, S= strong

**TABLE 7** Classification results from test set (2009 data set, 65 instances in total)

| Class Classification | I | W | M | S | Total | Classified Correctly (%) |
|---|---|---|---|---|---|---|
| I | 4 | 2 | 0 | 0 | 6 | 66.7% |
| W | 0 | 1 | 0 | 0 | 1 | 100.0% |
| M | 0 | 0 | 1 | 0 | 1 | 100.0% |
| S | 0 | 0 | 3 | 54 | 57 | 94.7% |
| Total | | | | | 65 | 92.3% |

I = insolvency, W = weak, M= moderate, S= strong

The results of applying the ordinal class classifier and DTL algorithms on the data introduced above depend on our selected financial ratios (attributes). The model shows a good performance and correctly classifies 98.7% from 10-fold cross-validation, 95.7% from 30% spilt test set, and 92.3% from the separated test set. The model can classify the minority class well but fail to recognize insolvency class in the separated test set (66.7% correctly classify). The relative importance of each attribute (input variable) is analyzed by calculating the weak class of the relationship between each input and output attribute.

**TABLE 8** Performance evaluation measure

| | Evaluation | |
|---|---|---|
| Cross-validation method | MAE | RMSE |
| 10 fold cross-validation | 0.0132 | 0.0838 |
| 30% spilt test set | 0.0281 | 0.1475 |
| Test set (2009 data set) | 0.0453 | 0.1985 |

MAE- Mean absolute error
RMSE- Root mean squared error

Table 8 presents performance evaluation measure of numeric prediction. In this study, we evaluated the performance of prediction by MAE and RMSE. The MAE and RMSE are given by

Mean absolute error (MAE)

$$= \frac{|p_1 - a_1| + .... + |p_n - a_n|}{n}$$

Root mean squared error (RMSE)

$$= \sqrt{\frac{(p_1 - a_1)^2 + ... + (p_n - a_n)^2}{n}}$$

Where, $P_1$ ,$P_2$.,.., $P_n$ denote the predicted values on the test instances and $a_1$ ,$a_2$.,.. ,$a_n$ denote the actual values.

## 5. Conclusions

From the experiment setting and results reported in the previous section, the results indicate that the obtained model can solve the problems of the multi-class classification and

also the imbalanced data set. In this study, we employ the ordinal class classifier to solve the multi-class problem, so that our model can classify the solvency condition of Thai Non-life insurance companies into four cases, strong, moderate, weak, and insolvency. To attack the problem of imbalanced data set, we use the standard resample technique which can highly improve the accuracy of the minority class which is the class that we are interested. Our final model are useful for insurance regulators, auditors, investors, management, policy holders, and related party to determine the priority for further examinations as an early warning system. In our further research, we will apply the ensemble methods and standard classifiers proposed here to better improve the imbalanced data set problem.

## References

[1] P. Georgios, An Overview of Solvency Supervision, Regulations and Insolvency prediction, Belgian Actuarial Journal, Vol. 8, 2009, pp.37-53.

[2] H. Scott, and N. Jack M., A Regression-Based Methodology for Solvency Surveillance in the Property-Liability Insurance Industry, The Journal of Risk and Insurance, Vol. 53, 1986, pp. 583-605

[3] T. James S., and P. George E., A Multivariate Model for Predicting Financially Distressed Property-Liability Insurance, The Journal of risk and Insurance, Vol.40, 1973, pp.327-338

[4] A. Jan Mills, and S. J. Allen, Using Best's Ratings, Financial ratio and prior probabilities in solvency prediction, The Journal of Risk and Insurance, Vol.55, 1988. pp. 229-244.

[5] C. James M., and H. Robert E., Life Insurer Financial Distress: Classification Models and Empirical Evidence, The Journal of Risk and Insurance, Vol.62, 1995, pp. 764-775.

[6] C. J. David , G. Martin F., and P. Richard D., Regulatory Solvency Prediction in Property-Liability Insurance: Risk-Based Capital, Audit Ratios, and Cash Flow Simulation, The Journal of Risk and Insurance, Vol.66, No.3, 1998, pp. 417-458.

[7] B. Ran, and H. Robert A., Classifying Financial Distress in the Life Insurance Industry, The Journal of Risk and Insurance, Vol.57, 1990, pp.110-136.

[8] A. Jan M., and C. Anne M., Using Best's Ratings in Life Insurer Insolvency Prediction, The Journal of Risk and Insurance, Vol. 61, 1994, pp. 317-327.

[9] L. Suk Hun, and U. Jorge L., Analysis and Prediction of Insolvency in the Property-Liability Insurance Industry: A Comparison of Logit and Hazard Models, Journal of Risk and Insurance, Vol.63, 1996, pp. 121-130.

[10] B. Ran, and H. John, The Merger or Insolvency Alternative in the Insurance Industry, The Journal of Risk and Insurance, Vol.64, 1997, pp. 89-113.

[11] E.H. Duett, and R.A. Hershbarger, Identifying Financial Distress in the Property-Casualty Industry, Journal of the Society of Insurance Research, Vol.21, 1990, pp. 33-45.

[12] C.S. Huang, R.E. Dorsey, and M.A. Boose, Life Insurer Financial Distress Prediction: A Neural Network Model, Journal of Insurance Regulation, Vol.13, 1994, pp. 131-167.

[13] B. Patrick L., C. William W., G. Linda L., and P. Utai, A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency, The Journal of Risk and Insurance, Vol. 61, 1994, pp. 402-424.

[14] K. Bert, N.E.W.S.: A model for the evaluation of non-life insurance companies, European Journal of Operational Research, Vol. 98, 1997, pp.419-430.

[15] H. Shu-Hua, and W. Thou-jen, A study of financial insolvency prediction model for life insurers. Expert Systems with Applications, Vol.36, 2009, pp.6100-6107.

[16] S.S. Sancho, F.V. Jose-Luise, S.V. Maria Jesus, B.C. Calos, Genetic programming for the prediction of insolvency in non-life insurance companies, Computers & Operations Research, Vol. 32, 2005, pp. 749-765.

[17] F. Eibe, and H. Mark, A simple approach to ordinal classification. In L. de Raedt, & P. A. Flach (Eds.), Proceedings of the Twelfth European Conference on Machine Learning, 2001, pp. 145–156.

[18] H. Mark, F. Eibe, H. Geoffrey, P. Bernhard, R. Peter, and W. Ian H., 'The WEKA Data Mining Software: An Update', SIGKDD Explorations Vol.11, Issue 1. 2009.

[19] J.R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., 1993.

[20] S. Yanmin., K. Mohamed S., W. Andrew KC., W. Yang, Cost-sensitive boosting for classification of imbalanced data Pattern Recognition, Vol.40 , 2007, pp. 3358-3378.

[21] V. Garcia, J.S. Sánchez, R.A. Mollineda, R. Alejo, J.M. Sotoca, The class imbalance problem in pattern classification and learning, 2007, pp. 283-291.

# Uniform Fiber Bragg Grating modeling and simulation used matrix transfer method

**Abdallah IKHLEF, Rachida HEDARA, Mohamed CHIKH-BLED**

**Laboratoire de Télécommunications, Département de Génie Electrique et d'Electronique**
**Faculté de Technologie, Université Abou-Bekr Belkaïd -Tlemcen**
**BP 230, Pôle Chetouane, 13000 Tlemcen- Algeria**

**Abstract**

This paper presents the modeling and simulation of an optical fiber Bragg grating for maximum reflectivity, minimum side lobe. Gating length represents as one of the critical parameters in contributing to a high performance fiber Bragg grating. The reflection spectra and side lobes strength were analyzed with different lengths .The side lobes have been suppressed using raised cosine apodization while maintaining the peak reflectivity. Such simulations are based on solving coupled mode equations by transfer matrix method.

*Keywords: Fiber Bragg grating, Reflection, Apodization, simulation Transfer Matrix Method.*

## 1. Introduction

Optical fiber gratings are important components in fiber communication and fiber sensing fields. For normal fiber gratings, by properly choosing the period, length, index modulation amplitude, chirp and apodization function, one can flexibly design and optimize grating reflection or transmission spectra to satisfy many applications [1]. Although optical fibers have been used for many decades, the last 10 to 20 years have shown a lot of further development. The introduction of FBGs, photonic crystal fibers and new plastic optical fibers, to name only the most important new fields, has dramatically widened the range of possible applications.

The FBGs are used extensively in telecommunication industry for dense wavelength division multiplexing, dispersion compensation [2,3], laser stabilization, and Erbium amplifier gain flattening, simultaneous compensation of fiber dispersion , dispersion slope and optical CDMA [4,5]. By exploiting the characteristics exhibited by these gratings, numerous areas have been marked in which their usage has brought drastic advancements and continues to do the same. The FBG works on the principle that when ultraviolet light (UV) illuminates a certain kind of optical fiber, the refractive index of the fiber is changed permanently, this effect is called photosensitivity. Alternatively, the refractive index will last for several years if it is followed by proper annealing [6].

The optical fiber with germanium doped core remains the most important material for grating purposes. The first in-fiber Bragg grating was demonstrated by Ken Hill in 1978 [7].

Initially, the gratings were fabricated using a visible laser propagating along the fiber core. In 1989, Gerald Meltz and colleagues demonstrated the much more flexible Transverse Holographic Technique [6] where the laser illumination came from the side of the fiber. This technique uses the interference pattern of ultraviolet laser light to create the periodic structure of FBGs. Since this discovery, the development in the field of Bragg gratings has experienced a tremendous growth. FBGs offer ample advantages but the most important is the flexibility in spectral characteristics. Many researchers have been work done in this field also [1, 8].
There are a number of parameters on which the spectra of FBG has shown dependency such as change in refractive index, bending of fiber , grating period, mode excitation conditions, temperature and fiber Bragg grating length [9,10,11,12].

In this paper, the effect on the Reflection spectra of FBG is analyzed at the varied grating length. The paper is divided into following sections. Section 2 covers the theory and modeling (coupled mode theory and transfer matrix method) of FBGs as well as the working principle of FBGs. Section 3 deals with the results and discussion about the modeling and simulation work done on FBGs at typical specifications using MATLAB. Lastly, section 4 draws the conclusion of the work done.

## 2. Theory

The propagation of light along a waveguide can be described in terms of a set of guided electromagnetic waves called the modes of waveguide. In optical fibers the core-cladding boundary conditions lead to coupling between the electric and magnetic field components.

Each mode has its specific propagation constants. If the periodic perturbation is introduced alongside the fiber the mode will exchange its power. This phenomenon is known as mode coupling. Fiber gratings can be broadly classified into two types: Bragg gratings (also called reflection and short-period gratings), in which coupling occurs between modes traveling in opposite directions; and transmission gratings (also called longperiod gratings), in which the coupling is between modes traveling in the same direction.

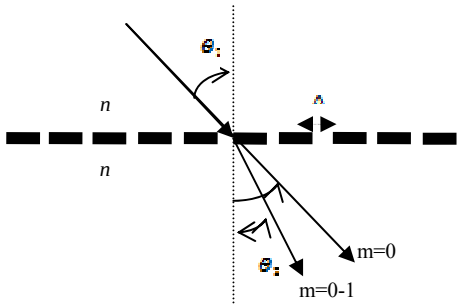



Fig. 1 the diffraction of light wave by a grating

A fiber grating is simply an optical diffraction grating, and thus its effect upon a light wave incident on the grating at an angle can be described by the familiar grating equation:

$$n\sin(\theta_2) = n\sin(\theta_1) + m\frac{\lambda}{\Lambda}. \qquad (1)$$

Where $\theta_2$the angle of the diffracted wave and the integer m is determines the diffraction order (Fig. 1). [11].

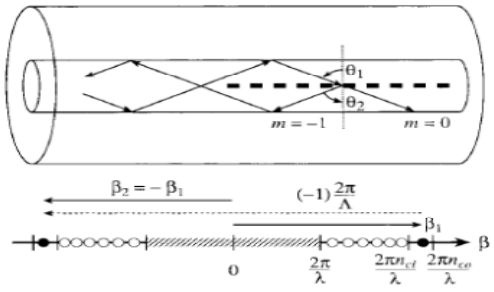



Fig. 2 Core mode Bragg reflection by a fiber Bragg grating

Fig. 2 illustrates reflection by a Bragg grating of a mode with a bounce angle of $\theta_1$ into the same mode traveling in the opposite direction with a bounce angle of $\theta_2$ =-$\theta_1$. $\beta$ is the $z$ component of wave propagation constant k is the main parameter in describing fiber modes, is simply:

$$\beta = \frac{2\pi}{\lambda}n_{eff}\text{ ,where }\quad n_{eff} = n_{co}\sin\theta \qquad (2)$$

The mode remains guided as long as $\beta$ satisfies the condition $n_2 k < \beta < n_1 k$

Where $n_1$ and $n_2$ are core and cladding refractive index and:

$$k = \frac{2\pi}{\lambda} \qquad (3)$$

## 2.1 Coupled mode theory

Coupled Mode Theory is a method to analyze the light propagation in perturbed or weakly coupled waveguides. The basic idea of the *Coupled Mode Theory method* is that the modes of the unperturbed or uncoupled structures are defined and solved first. Then, a linear combination of these modes is used as a trial solution to Maxwell's equations for complicated perturbed or coupled structures. After that, the derived coupled mode equations can be solved analytically or by numerical methods.
The coupled-mode equations describe their complex amplitudes, *A(z)* and *B(z):*

$$\frac{dA(z)}{dz} = jB(z)K\ exp[j(\beta_1 - \beta_2)z]$$
$$\frac{dB(z)}{dz} = jA(z)K^*\ exp[-j(\beta_1 - \beta_2)z] \qquad (4)$$

In our simulation, the coupled mode equations are based on non-orthogonal coupled mode theory [13, 14]. Both the waveguide nature coupling and grating coupling are considered. In order to formulate the coupled mode equations, waveguide modal constants, fields, and coupling coefficients are calculated based on waveguide and grating profiles.
In our work the well-known transfer matrix method is applied to solve the couple mode equations and to obtain the spectral response of the fiber grating. In this approach, the grating is divided into uniform sections; each section is represented by a 2*x*2 matrix. By multiplying these matrices, a global matrix that describes the whole grating is obtained.

## 2.2 Transfer matrix

Divide the grating into a sufficient number $N$ of sections so that each section can be approximately treated as uniform. Let the section length be $\Delta = L/N$ .By applying the appropriate boundary conditions and solving the coupled-mode equations similar to the procedure in Section (2.1), we find the following transfer matrix relation between the fields at z and at z + Δ.

$$\begin{bmatrix} u(z+\Delta) \\ v(z+\Delta) \end{bmatrix} = \begin{bmatrix} \cosh(\gamma\Delta) - i\frac{\Delta\beta}{\gamma}\sinh(\gamma\Delta) & -\frac{\kappa}{\gamma}\sinh(\gamma\Delta) \\ i\frac{\kappa}{\gamma}\sinh(\gamma\Delta) & \cosh(\gamma\Delta) + i\frac{\Delta\beta}{\gamma}\sinh(\gamma\Delta) \end{bmatrix} \begin{bmatrix} u(z) \\ v(z) \end{bmatrix} \quad (5)$$

We can connect the fields at the two ends of the grating through

$$\begin{bmatrix} u(L) \\ v(L) \end{bmatrix} = T \begin{bmatrix} u(0) \\ v(0) \end{bmatrix} \quad (6)$$

Where

$$T = T_N * T_{N-1} * \ldots . T_i \ldots . T_1 \quad (7)$$

The matrix $T_j$ is the transfer matrix written in (5) with $\kappa$ the coupling coefficient of the jth section. As a result, T is a 2 × 2 matrix with elements

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \quad (8)$$

The matrix $T_i$ for one section is defined by

$$T_i = \begin{bmatrix} \cosh(\gamma l_i) - i\frac{\Delta\beta}{\gamma}\sinh(\gamma l_i) & -\frac{\kappa}{\gamma}\sinh(\gamma l_i) \\ i\frac{\kappa}{\gamma}\sinh(\gamma l_i) & \cosh(\gamma l_i) + i\frac{\Delta\beta}{\gamma}\sinh(\gamma l_i) \end{bmatrix} \quad (9)$$

The reflection coefficient is calculated by the relation:

$$R = \frac{T_{21}}{T_{11}} \quad (10)$$

The characteristics response from Bragg Grating can be analyzed fully described by

1. The center wavelength of Grating $\lambda_B$
2. Peak reflectivity $R_{max}$ of grating which occur at $\lambda_B$
3. Physical length of Grating *L*.

For a grating with uniform index modulation and period the reflectivity is given by

$$R(L,\lambda) = \frac{\kappa^2 sinh^2(\gamma L)}{\Delta\beta^2 sinh^2(\gamma L) + \kappa^2 cosh^2(\gamma L)} \quad (11)$$

R: Grating reflectivity as a function of both grating length and wavelength.

*L*: total length of grating.

The coupling coefficient κ (z) is defined by the following equation

$$\kappa(z) = \frac{\pi}{\lambda}.\Delta n.g(z).v \quad (12)$$

If the FBG is uniform, then $\Delta n_{eff}$ is constant, g (z) =1,*v* the fringe visibility and is usually estimated at 1.

$$\kappa = \frac{\pi \Delta n_{eff}}{\lambda}.$$

$\Delta\beta$: wave vector detuning, given by $\Delta\beta = \beta - \frac{\pi}{\Lambda}$

$\beta$: Fiber core propagation constant, given by $\beta = \frac{2\pi n_0}{\lambda}$

$$\gamma = \sqrt{\kappa^2 - \Delta\beta^2}.$$

For light at the Bragg grating center wavelength,$\lambda_B$, there is no wave vector detuning and so $\Delta\beta = 0$. The reflectivity function then becomes

$$\mathrm{R(L},\lambda) = tanh^2(\gamma L) \quad (13)$$

## 3. RESULTS AND DISCUSSION

The parameters used for Simulation are core index = 1.47, cladding index = 1.457, λ =1550 nm, change in refractive index, $\Delta n_{eff} = 1e^{-4}$, grating period, Λ= 5.3e-7. The grating length has been varied from L =05 mm to 40mm.

For different values of grating length (table 1), Reflection spectra was obtained and analyzed. From the spectra, it was confirmed that the spectral properties of uniform gratings comes out to be similar to *sinc* function. The reflection spectra for different grating length 5 mm, 07 mm, 10 mm, 15 mm and 25 mm is shown below in (fig. 3,4,5,6,7).At L=05mm, 07mm, 10mm, and 15 mm successively the maximum reflectivity is 59.86%**,** 79.04%**,** 93.28%**,** 99.09% and 99.98%.At L=25 mm, the reflectivity reaches 99.98% but increase in the reflectivity of sides lobes .After that, if the length is incremented further, it is observed that maximum reflectivity maintains the same value of 99.99%.

Table 1: Reflectivity for different grating lengths

| Grating Length (mm) | Reflectivity obtained (%) |
|---|---|
| 05 | 59.86 |
| 07 | 79.04 |
| 10 | 93.28 |
| 15 | 99.09 |
| 16 | 99.39 |
| 18 | 99.73 |
| 20 | 99.90 |
| 22 | 99.95 |
| **25** | **99.98** |
| 28 | 99.99 |
| 30 | 99.99 |
| 35 | 99.99 |
| 40 | 99.99 |

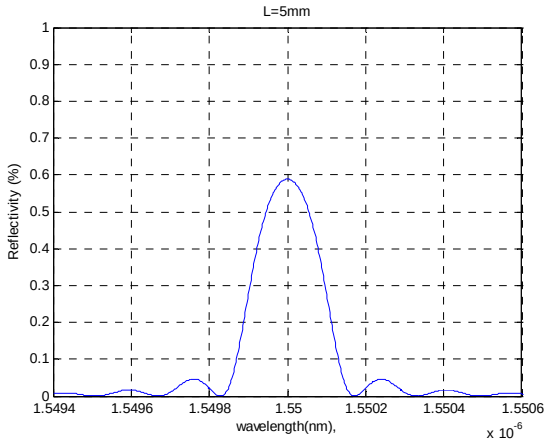



Fig. 3 Reflection spectrum at L = 05mm

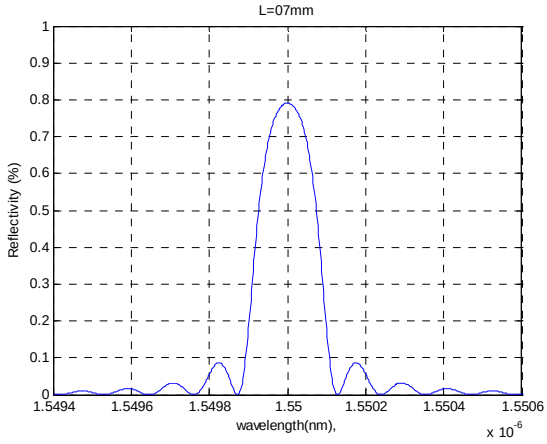



Fig. 4 Reflection spectrum at L = 07mm

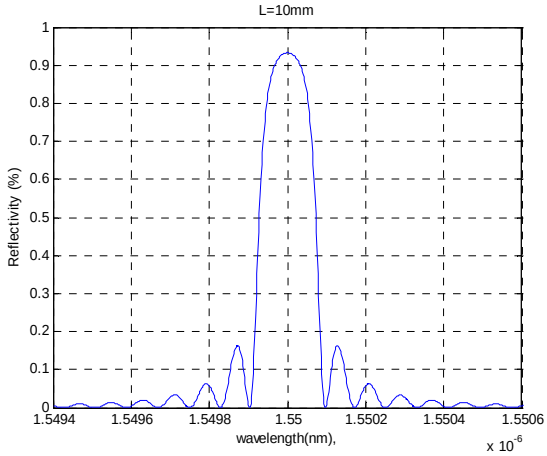



Fig. 5 Reflection spectrum at L = 10mm

The reflectivity reaches 99.99% but it is accompanied with a significant increase in the reflectivity of side lobes.

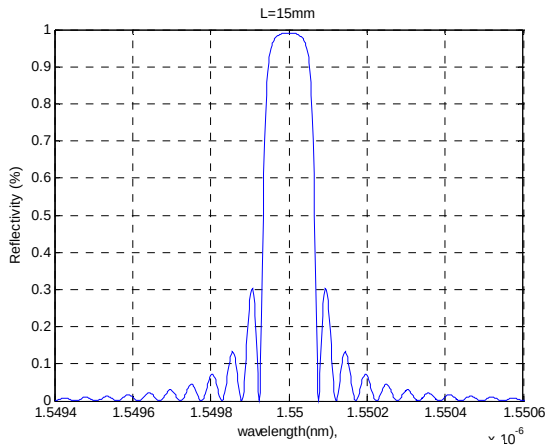



Fig. 6 Reflection spectrum at L = 15mm

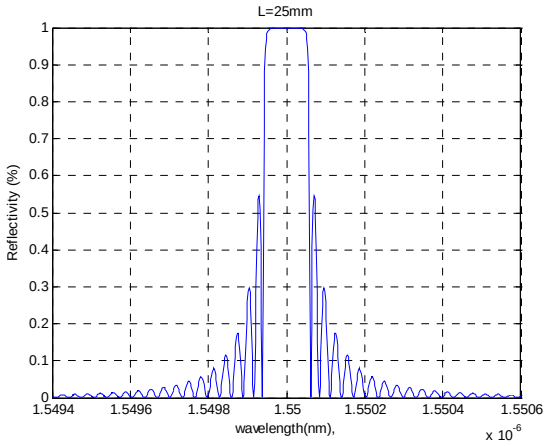



Fig. 7 Reflection spectrum at L = 25mm

As shown in Fig. 8 from the above results, upon consideration of the reflectivity of the uniform FBG, it was confirmed that the simulated uniform FBG showed better performance as the grating length increased and achieved 99.98 % reflection at the grating length of 25 mm.

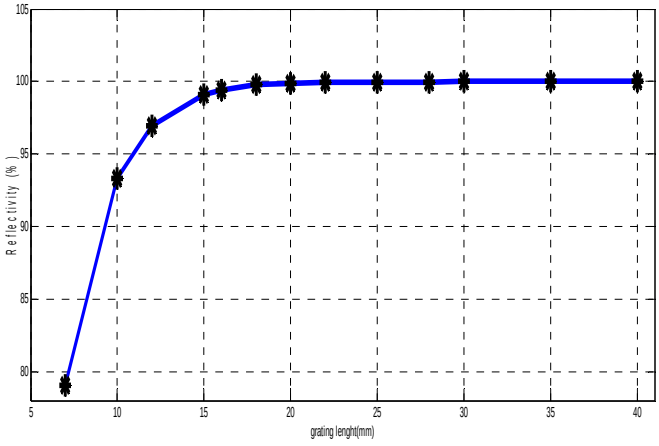



Fig. 8 relation between uniform FBG reflectivity and grating length

A uniform FBG is accompanied by a series of side-lobes adjacent to the Bragg wavelength. A very effective method for eliminating the side-lobes of an FBG is apodization.

Apodization is achieved by a contoured inscription of the grating in order to reduce the refractive index change towards the ends of the grating.

The transfer matrix method can be adjusted to accommodate for the apodization of an FBG by replacing $g(z)$ in equation (12) with for example, the following raised cosine apodization.

$$g(z) = \Delta n_{eff}.\frac{1}{2}.\left\{1 + \cos(\frac{z\pi}{L}\right\}. \qquad (14)$$

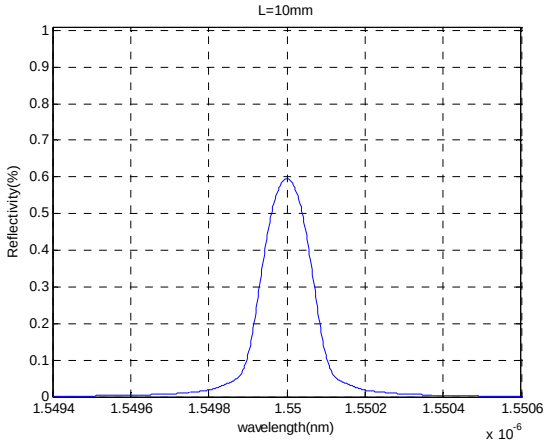



Fig. 9 Apodized reflectance spectrum at L = 10mm

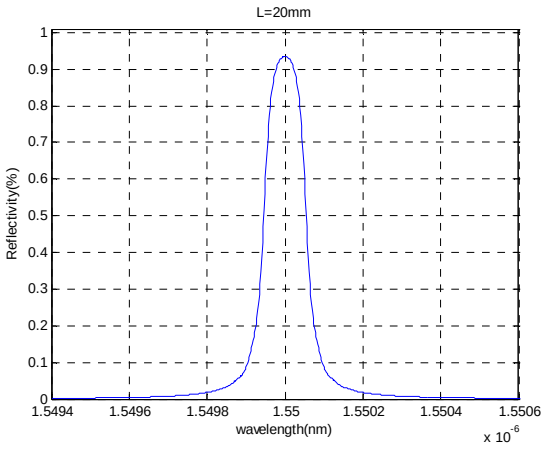



Fig. 10 Apodized reflectance spectrum at L = 20mm

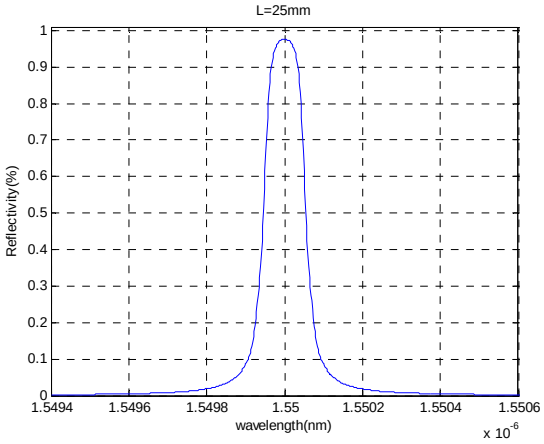



Fig. 11 Apodized reflectance spectrum at L = 25mm

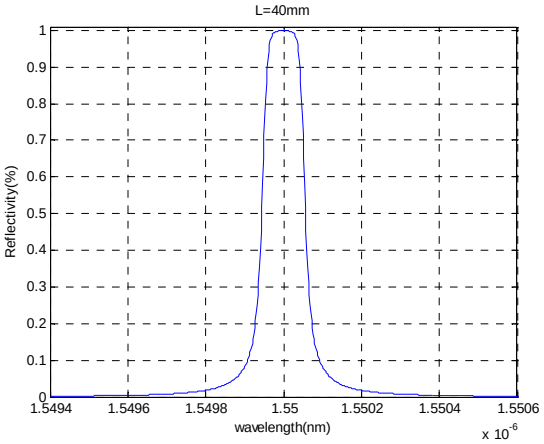



Fig. 12 Apodized reflectance spectrum at L = 40mm

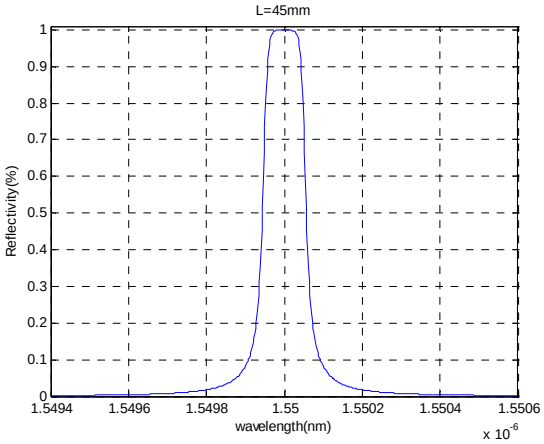



Fig. 13 Apodized reflectance spectrum at L = 45mm

Other apodization functions that are used in the communications industry include pure cosine, Gaussian, sinc and Kaiser profils [11, 15].

Fig. 9, 10,11,12,13 illustrates the reflectance spectrum response of an apodized FBG for different grating length. At L=10mm,20mm,25mm,40mm and 45mm the maximum reflectivity is is 59.47%**,** 93.55%**,** 97.62%**,** 99.88% and 99.99%.Note that all of side lobes have been completely eliminated But reflected power can be increased by increasing the length of apodized FBG.

Table 2: Reflectivity of Apodized FBG for different grating lengths

| Grating Length (mm) | Reflectivity obtained (%) |
|---|---|
| 10 | 59.47 |
| 15 | 82.94 |
| 20 | 93.55 |
| 25 | 97.62 |
| 28 | 98.70 |
| 30 | 99.14 |
| 35 | 99.69 |
| 40 | 99.88 |
| 42 | 99.92 |
| **45** | **99.99** |
| 50 | 99.99 |
| 55 | 99.99 |
| 60 | 99.99 |

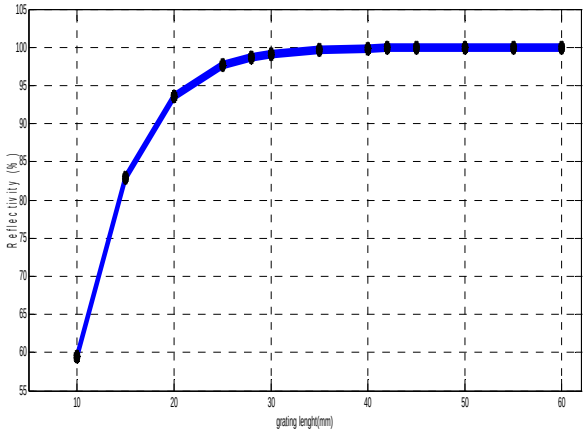



Fig. 14 relation between Apodized F BG reflectivity and grating length

The reflectivity showed an exponential increase over the elevation of grating lengths, as shown in Fig. 14. From the above results, upon consideration of the reflectivity elevation of apodized FBG, it was confirmed that the simulated apodized FBG showed better performance as the grating length increased and achieved 99.99 % reflection at the grating length of 45 mm.

Based on the results obtained for uniform FBG, the variation in the reflectivity, the side lobe suppression and apodized FBG, optical fiber Bragg grating for maximum reflectivity, minimum side lobe is tabulated below:

| | Length FBG | reflectivity | Sides lobe |
|---|---|---|---|
| Unifom FBG | 10 mm | 99,99% | Between 10% and 50% |
| Apodized FBG | 45 mm | 99;99% | 0% |

## 4. Conclusion

In this work, we have described the signal characteristics of FBG with various grating lengths using simulation method. We conducted quantitative analyses on the reflectivity with the increases of grating length. The conclusions obtained from this study are as follows.

1. The reflectivity of fiber grating increases with the increase in grating length.
2. The reflectivity increased with the elevation of grating length in which it achieves 99,99% in reflection at grating length 10 mm and maintained this value for longer length.
3.We got full Suppression of side lobes in reflectivity curve for Raised cosine Apodization at the cost of reduced reflected power. But reflected power can be increased by increasing the length of apodized FBG.The reflectivity increased with the elevation of grating length in which it achieves 99,99% in reflection at grating length 45 mm and maintained this value for longer length.

## References

[1] Erdogan, Turan , " Fiber Grating Spectra", Journal of Lightwave Technology, Vol. 15, No. 8, August 1997, pp. 1277 – 1294

[2] Zhongwei Tan and al, "Transmission system over 3000km with dispersion compensated by chirped fiber Bragg gratings», Optical communication (2007)

[3] I . Navruz , N. Fatma Guler, "A novel technique for optical dense comb filters using sampled fiber Bragg gratings ",Optical Fiber Technology 14 (2008) 114–118

[4] Qiang Wu, Chongxiu YuKuiru Wang, Xu Wang, Zhihui Yu, H.P. Chan,Pak L. Chu,"New Sampling-Based design of Simultaneous Compensation of Both Dispersion and Dispersion Slope for Multichannel Fiber Bragg Gratings", IEEE Photonics Technology Letters, Vol. 17,No. 2, Febuary 2005, pp.381-383

[5] J. Magné, D.-P. Wei, S. Ayotte, L. A. Rusch and S. LaRochelle " Experimental Demonstration of Frequency-

Encoded Optical CDMA using Superimposed Fiber Bragg Gratings" OSA,fiber optical communication ,(2003)
[6] Kashyap, R: Fiber Bragg Gratings, San Diego, Academic Press, 1999, ISBN 0-12-400560-8
[7] O.H. Hill, G. Meltz: "Fiber Bragg Grating Technology Fundamentals and Overview", Journal of Lightwave Technology 15(8) (1997), pp. 1263-1276
[8] T. Mizunami, T. V. Djambova, T. Niiho, and S. Gupta, "Bragg gratings in multimode and few-mode optical fibers," J. Lightwave Technol., vol. 18, Feb. 2000, pp. 230–234
[9] Ho Sze Phing, Jalil Ali, Rosly Abdul Rahman and Bashir Ahmed Tahir:" Fiber Bragg grating modeling, simulation and characteristics with different grating lengths",Journal of Fundamental Sciences, July 2007
[10] Ravijot Kaur, Manjit Singh Bhamrah"Effect of Grating length on Reflection Spectra of Uniform Fiber Bragg Gratings" International Journal of Information and Telecommunication Technology, Vol. 3, No. 2, 2011 ISSN (Online): 0976-5972
[11] Sunita Ugale et. al. "Fiber Bragg Grating Modeling, Characterization and Optimization with different index profiles", International Journal of Engineering Science and Technology Vol. 2(9), 2010, 4463-4468
[12] G. GOPALAKRISHNANand al, "Fiber Bragg Grating based temperature and strain sensor simulation for biomedical applications», OPTOELECTRONICS AND ADVANCED MATERIALS Vol. 2, No. 1, January 2008, p. 10 – 14
[13] A. Herma and W. Hugang, «Coupled-Mode Theory", IEEE, Proceedings Of the IEEE, Vol. 19, pp. 1505-1518, 1991.
[14] H. A. Haus, W. P. Huang, S. Kawakami, and N. A. Whitaker, "Coupled-Mode Theory of Optical Waveguides", IEEE, Journ. of Ligth. Tech., Vol. 05, pp. 16-23, 1987.
[15] B.B.Padhy et al. "optimization of intrgrating sensing using fiber Bragg grating" International Journal of Engineering Science and Technology, Vol. 2(9), 2010, 4463-4468

# Valuable Internet Advertising and Customer Satisfaction Cycle (VIACSC)

**Muhammad Awais (Assistant Professor)**
**Department of Computer Science, NFC Institute of Engineering & Fertilizer Research**
**Faisalabad, Pakistan, 38000**

**Tanzila Samin (Lecturer)**
**School of Business Management, NFC Institute of Engineering &Fertilizer Research**
**Faisalabad, Pakistan, 38000**

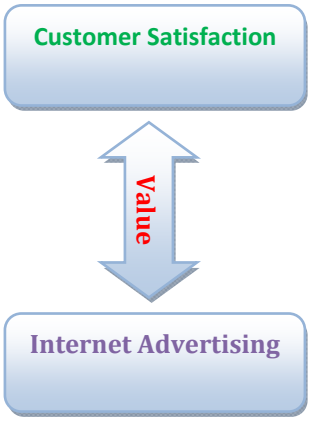
**Muhammad Bilal**
**Faisalabad, Pakistan, 38000**
**University of Engineering & Technology Lahore, Pakistan.**

***Abstract***

Now-a-days it is very important for the business persons to attract their target customers towards their products through valuable mode of promotion and communication. Increasing use of World Wide Web has completely changed the scenario of business sector. Customized products and services, customers preferences, @ and dot com craze have elevated the importance of internet advertising. This research paper investigates valuable internet advertising which will help to enhance the value of internet advertising. In this research we have compared internet advertising with television advertising and found advertising will survive and grow if it focuses on being valuable. This research paper concentrates that the business objectives can be achieved if it has strong attractive, informative and valuable internet advertising. The basic purpose of VIACSC is to identify the latent need of the customer through advertisement and inform him/her about a product which will help to retain business image, customer satisfaction and then loyalty. We have introduced certain steps in valuable internet advertising, and then customer satisfaction cycle through valuable internet advertising and finally 5A's which enhance further customers attraction towards the internet advertising.

***Key words:*** *Marketing Mix, Promotion Mix, Valuable Internet Advertising and Customer Satisfaction Cycle (VIACSC)*

## 1. Introduction




Internet advertising is developed to promote information about products and services to target customers in an optimistic approach of making them agree to buy products and services. Now- a- days it is rarely possible for a business firm to become successful without involving in advertising efforts specifically valuable internet advertising in which customers can easily search about their required products and services because it saves their time and money. Valuable internet advertising is a key factor in the success of any product and service, and in this competitive era companies are ready to invest whole heartedly in their internet advertising campaign not only to make their products and services successful but also to achieve customer retention. Generally internet advertising is working to achieve four goals ***(1)*** detail but to the point information about the products and services ***(2)*** increase sales ***(3)*** business

image, and ***(4)*** customer retention. In simpler words effective internet advertising attempts to inform, convince and retain the customers. The major goal of a business is to bring out the latent needs of the customers through their products and services by advertising them in an effective manner and to make the customer ready to buy them. Once the business reputation is established, the products and services are positioned in the market; company willingly keeps on investing on advertising to persuade customers every time differently to buy their products and services. According to business directory customer satisfaction means "The degree of satisfaction provided by goods and services of a company as measured by the number of repeating customers "Customer satisfaction differs depending on the situation and the product or service. A customer may be satisfied with the product or service, an experience, a purchase decision, a sales advertisement, a web site or any attribute of any of these. Customer satisfaction is a highly personal assessment that is greatly influenced by individual expectations. Some definitions are based on the observation that customer satisfaction or dissatisfaction results from either the confirmation or disconfirmation of individual expectations regarding a service or product. To avoid difficulties stemming from the kaleidoscope of customer expectations and differences, some experts urge companies to "concentrate on a goal that's more closely linked to customer equity" Instead of asking whether customers are satisfied, they encourage companies to determine how customers hold them accountable. At the same time let us see the value which plays an intermediary role between internet advertising and customer satisfaction and it stands for the qualitative goals which a company wants to achieve through internet advertising by providing customer satisfaction and at the end company and business seek to grow or maintain their performance.

## 2. Literature Review

Nothing attracts people and businesses more than money. These spent dollars are the driving force for advertising online. The philosophy is that companies will be able to target audiences with pinpoint accuracy. It will allow companies to track who is seeing their advertising? What action will be taken after they see it? And in some cases, tailor the advertising to the consumer. (Savitz, 1999). In order to make money businesses need to establish themselves with consumers. Moreover, to get consumers, businesses must advertise. "Old-line merchants no longer insist, as they did not long ago, that they can do without an Internet presence. Everyone is coming online" (Offline stores are moving, 1999). Advertising is a way of life and maybe the only life for some one. If the WWW were thought of as broadcast television, then 100 percent of the revenue would come from advertising (Zeff and Aronson, 1997). Along with this colorful landscape, came the world of electronic commerce (EC) which is the transaction online. Evidence from market research firms, such as Odyssey, suggested that more American households are making purchases over the Internet (Lohr, 1999). Hyland (1998) had briefly described how the Internet had become an accepted Communication medium in just five short years? A case study had compared the Internet to television. Everett-Thorp (1997) had detailed how being online has not changed the basic concept of advertising, but had put a whole new spin on it. Her article had stated why it is important to advertise online? Where the audience was? And how the competition for advertisements was getting fiercer all the time? Sterne (1997) was one of the best resources for this topic. He had explored the world of online advertising in depth. In some respects, it had gone further than the requirements set forth in this project. Tedeschi (1998) had briefly described direct marketing, which was advertising by email. His article had detailed different case studies of companies who used this method and had success. He also had talked about how this advertising model was used in the early Internet day? Hofacker (1999) had briefly described classified advertisements and Oikle (1997) had detailed newsgroups and Newsletters. Hofacker (1999) had treated advertising in a different way. He had detailed Internet marketing and how to create a Web site? How it should look? And the communication aspect of it.Oikle (1997) had detailed the power of email marketing with newsgroups and newsletters.

And now take a look to customer satisfaction Despite extensive research in the years since Cardozo's (1965) classic article, researchers have yet to develop a consensual definition of consumer satisfaction. Oliver (1997) addresses this definitional issue by paraphrasing the emotion literature, noting that "everyone knows what [satisfaction] is until asked to give a definition. Then it seems, nobody knows" (p. 13). Based on the perception that satisfaction has been defined, most research focuses on testing models of consumer satisfaction (e.g., Mano and Oliver 1993; Oliver 1993;while definitional considerations have received little attention. As Peterson and Wilson (1992) suggest, "Studies of customer satisfaction are perhaps best characterized by their lack of definitional and methodological standardization" (p. 62).

## 3. Promotion Mix

Promotion mix is an important tool in marketing and is used to create a message and information about a product and to disseminate it. And it can be done only through advertising and internet advertising. In advertising we include print media (news papers, magazines, directories, yellow pages, etc), electronic media (radio, television, etc) and now online or internet advertising (www- world wide web, banners, e-mail, skyscrapers, mini sites and pop ups) which is more effective, efficient, valuable, cost effective and time saving; and it can communicate the message to a large mass. Here we will discuss two important tools of promotion mix and they are internet advertising and television advertising.

## 4. Internet Advertising vs Television Advertising

In this comparison we will discuss various factors/parameters and try to prove how internet advertising is better than television advertising.

Let us compare both advertising techniques:

| Factors/ Parameters | Internet Advertising | Television Advertising |
|---|---|---|
| Time Constraint | Products and services can be browsed at any time | Specific telecast/broad cast timing |
| Accessibility | 24 hours/7days a week access | Limited accessibility |
| Features | Variety of features | Limited features |
| Price Comparison | Prices can be compared | Usually price is not mentioned |
| Comparison of features | Features can be compared | No features comparison can be made |
| Feedback | Quick feedback of the customer | Feedback cannot be quick |
| Privacy | Customer privacy policy | Privacy cannot be maintained |
| Choices | Multiple choices | Limited choices |
| Review by customer | Can be checked and browsed repeatedly | Cannot be rewind by the customer |

**Table 1**

In the above table it is mentioned that through internet advertising we can browse any product and service while sitting anywhere, any time and even repeatedly we can check the required information. While in television advertising specific timings are for telecasting, we cannot see the product according to our own wish and demand, we will have to wait for that to be advertised because here we cannot rewind the transmission. Another important point in internet advertising is that variety of features of the products are available, we can compare products of different brands at the same time and even their prices can be compared and decision can be made according to one's own affordability. For every type of business customer feedback does matter a lot, through internet customer give quick feedback where as in television we have to wait for a long time for customer's feedback. With the help of above comparison we can say that internet advertising is more valuable, effective, efficient, innovative, time saving and cost effective than advertising through television.

## 5. Steps in developing Valuable Internet Advertising (VIA)

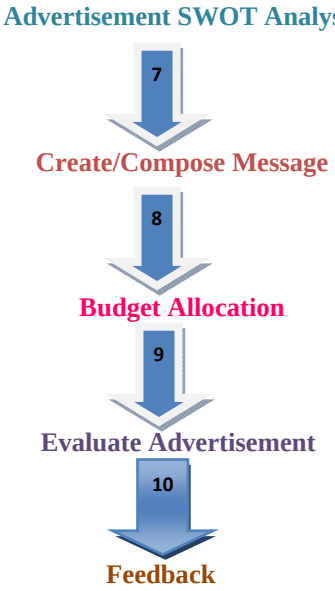



**Fig 1**

### 5.1 Identify latent need:

Latent needs are hidden requirements of humans. These are to be identified in order to create products that the customers don't even know they desire or, in some cases, solutions that customers have difficulty envisioning due to lack of familiarity with the possibilities offered by new technologies or because locked in an old mindset.

### 5.2 Focus target audience:

The target audience is a specific group of people within the target market at which the marketing message is aimed. Target Audiences are formed from different groups, for example: Adults, teens, children, mid-teens, pre-schoolers.

### 5.3 Hit potential market:

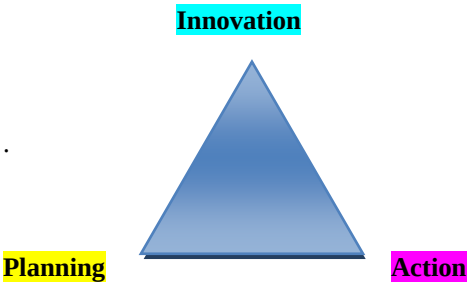
It is essential to become familiar with potential market; their habits, behaviors, likes, and dislikes. Markets differ in size, assortment, geographic scale, locality, types of communities, and in the different types of merchandise sold.

### 5.4 Define advertising strategy:

An advertising strategy is a campaign developed to communicate ideas about products and services to prospect customers for convincing them to buy the product.

### 5.5 Advertising goals/objectives:

Advertising must be goal oriented but for every advertisement three elements are important




### 5.6 Advertisement SWOT analysis:

The complete evaluation of an advertisement's strength, weakness, opportunities and threats is called SWOT analysis.

### 5.7 Create/Compose Message:

Before creating a message some points must be considered.

- What is the position of the product in the market?
- Who is our target audience?
- What is their cultural background?
- Either message will be based on serious thinking, humor, romance, and/or emotion?
- Which mode of communication is to be chosen?

### 5.8 Budget allocation:

A businessman knows better that he/she has a certain amount of money for advertisement and allocated budget will tend to dictate what advertisement is to be developed?

### 5.9 Evaluate advertisement:

Finally evaluate by rapid prototyping so it should not hurt anyone by religious, cultural, social, moral and ethical point of view.

### 5.10 Feedback:

Customer feedback is the best judge. Their feedback and response will give actual results of effectiveness.

## 6. Five A's in VALUABLE INTERNET ADVERTISING (VIA)

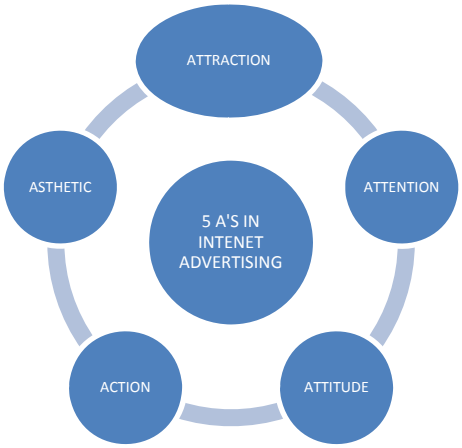



**Fig 2**

The above diagram shows advertising can be effective only if it touches the aesthetic sense of the customer which will attract him/her towards the product or service and the attention will be gained by the advertisement on internet. Because of the attention the customer will feel a change in his/her attitude and will think about the product or service, and his/her action to purchase the product will prove that advertisement is effective because customer has given his/her feedback through the purchase.

## 7. VALUABLE INTERNET ADVERTISING AND CUSTOMER SATISFACTION CYCLE (VIACSC)

Every human being has needs, wants and desires and seldom reaches at complete satisfaction level except for a short time. As one desire is satisfied, another pops up to take its place. When this is satisfied, still another comes into the foreground. It is the nature of human being that he/she practically always desires something new.

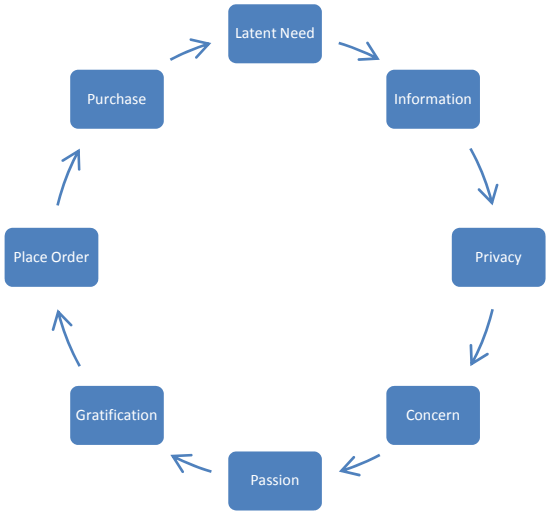



**Fig 3**

In Fig 3 we introduced VIACSC that deals with different phases of effectiveness of Internet Advertising as there is always a latent need of customer about a product/service, about which he/she wants to gather information but is not able due to many hindrances. The customer is privacy conscious as he/she come to know about a product or service through internet advertising and is ready to browse because privacy will not be disturbed now the concern will increase about gathering information and willingly search starts by making comparison of required product or service. While gathering information about different products he/she will be passionate and at the time of placing order, gratification level will be achieved and will place order, and finally purchase; then again a latent need will arise and customer will start working on it and the same cycle continues.

### Conclusion:

This research paper concludes that internet advertising becomes valuable when identifies customer's latent need, and deal with customer issue of privacy seems to be the main proponent in driving a new advertising concept. Apart from all above discussion main point of valuable advertising is to boost the product or service by identifying customer need and to satisfy them. By comparing two electronic mode of advertising we concluded that

internet is better form of advertising a product or service as compare to television.

**References:**

[1]Cardozo, Richard N. 1965. "An Experimental Study of Consumer Effort, Expectation and Satisfaction." Journal of Marketing Research 2 (August): 244-249.

[2] Everett-Thorp, K. (1997, December 16). Web advertising secrets. *CNet*. [Online].
Available:
http://www.builder.com/Business/Advertising

[3] Hofacker, C. F. (1999). *Internet Marketing*. Dripping Springs: Digital Springs, Inc.

[4] Hyland, T. (1998, November 19). Why internet advertising. *Internet Advertising Bureau (IAB)*. [Online]. Available
http://www.iab.net/advertise/content/adcontent.html.

[5] Lohr, S. (1999, March 22). Survey suggests consumers are taking to e-commerce.
*The New York Times on the Web*. [Online].
Available: http://www.nytimes.com.
[1999, March 22].

[6] Mano, Haim and Richard L. Oliver. 1993. "Assessing the Dimensionality and Structure of the Consumption Experience: Evaluation, Feeling, and Satisfaction." Journal of Consumer Research 20 (December): 451-466.

[7] Oikle, J. (1997). The power of e-mail marketing. *Intrepid Net Marketing*. [Online].
Available:
http://www.intrepidmarketing.com/Articles/emailmarketing.html

[8]Oliver, Richard L. 1993. "Cognitive, Affective, and Attribute Bases of the Satisfaction Response." Journal of Consumer Research 20 (December), 418-430.

[9]Oliver, Richard L. 1997. Satisfaction: A Behavioral Perspective on the Consumer. New York: The McGraw-Hill Companies, Inc.

[10]Peterson, Robert A. and William R. Wilson. 1992. "Measuring Customer Satisfaction: Fact and Artifact." Journal of the Academy of Marketing Science 20 (Winter): 61-71.

[11] Savitz, E. (1999, February 15). Web advertisers search for the promised land.
*The Industry Standard*. [Online]. Available:

[12] Sterne, J. (1997). *Advertising on the Web*. Que Education & Training.

[13] Tedeschi, B. (1998, December 8). Marketing by e-mail: Sales tool or spam?
*E-Commerce Report*. [Online]. Available: http://www.nytimes.com.

[14] Zeff, R. L., and Aronson, B. (1997). *Advertising on line on the Internet*. John Wiley & Sons

[15]WWW.BusinessDirectory.com

[16] http://wps.prenhall.com/bp_kotler_mm_13/

[17]http://openlibrary.org/books/OL9291253M/Advertising_Principles_and_Practice_(7th_Edition)_(Advertising_Principles_and_Practice)

[18]http://www.enotes.com/small-business-encyclopedia/advertising-strategy

[19] WWW.Learnmarketing.com

[20]http://www.worldata.com/wdnet7/articles/the_history_of_Internet_Advertising.htm

# Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification

**A. K. Santra[1], S. Jayasudha[2]**

[1] **Dean, CARE School of Computer Applications , Trichy – 620 009, India**.

[2] **Research Scholar, Bharathiar University, Coimbatore – 638401, India.**

**Abstract**

Web Usage Mining (WUM) is the process of extracting knowledge from Web user's access data by exploiting Data Mining technologies. It can be used for different purposes such as personalization, system improvement and site modification. Study of interested web users, provides valuable information for web designer to quickly respond to their individual needs.

The main objective of this paper is to study the behavior of the interested users instead of spending time in overall behavior. The existing model used enhanced version of decision tree algorithm C4.5. In this paper, we propose to use the Naive Bayesian Classification algorithm for classifying the interested users and also we present a comparison study of using enhanced version of decision tree algorithm C4.5 and Naive Bayesian Classification algorithm for identifying interested users. The performance of this algorithm is measured for web log data with session based timing, page visits, repeated user profiling, and page depth to the site length. Experimental results conducted shows that the performance metric i.e., time taken and memory to classify the web log files are more efficient when compared to existing C4.5 algorithm.

**Keywords:** Web Usage Mining, Web Mining, Web Log Files, Classification

## 1. Introduction

The ways in which users interact with a World Wide Web (Web) site provide enormous data processing on the usefulness and effectiveness of Web design elements and content built in it. Yet the informative log files recorded by Web servers, and client logs, offer potentially useful data about users Web site interactions. These data may be segregated and studied to generate inferences about Web site design, to test prototypes of Web sites or their modifications over time, and to test theoretical hypotheses about the effects of different design variables on Web user behavior.

Web usage mining involves with the application of data mining methods to discover user access patterns from web data. The main task of web usage data is to capture web-browsing behavior of users from a specified web site. Web usage mining can be classified according to kinds of usage data examined. In our context, the usage data is web log data, which maintains the information regarding the user navigation. Our work concentrates on web usage mining.

This paper proposes classification of web log data and studying the interested users from them. Due to the involvement of uninterested users in the web log, the original log cannot be used as a process in the web usage mining procedure. Thus in the first phase the web log data is preprocessed, to extract the interested data and then to proceed with the extracted data. During this phase, the actual size of the database will be minimized to certain extent. The second phase consists of segregating the data using Naive Bayesian Classification. The remainder of the paper is organized as follows. In section 2, we discuss the related work. In section 3, Naive Bayesian Classification approach is discussed in detail. Algorithm and mathematical evaluation is explained in section 4. Results on the experiments conducted are discussed in section 5. Finally conclusions is discussed in section 6.

## 2. Related Work

Classification of web log data using naïve Bayesian method is one of the well-known approaches that improve the overall performance of the web server. In this section, we provide taxonomy regarding web mining, classification rule mining methodology based on decision trees, the algorithm C4.5 that have been used in the existing work to identify the interested users.

Jie Zhang and Ali., A. Ghorbani [1] proposed Web usage mining plays an important role in the personalization of Web services. Users' access to pages of the Website should be separated into user sessions. The required user sessions are extracted from the Web server log. Several approaches have been proposed. In this paper we consider two different approaches in initially defining Web mining. First was a 'process-centric view', that defined Web mining as a sequence of tasks. Second was a 'data-centric view', which defined Web mining in proportion to the types of Web data that was being used in the mining process. Mahesh Thylore Ramakrishna1, Latha Kolal Gowdar, Malatesh Somashekar Havanur, Banur Puttappa Mallikarjuna Swamy [2] these authors follow the data-centric view, and refine the definition of Web mining.

Alka Gangrade, Durgesh Kumar Mishra, Ravindra Patel [3] focused on the review of the techniques for privacy preserving classification under multi-party environment. Further, the two approaches, the classification model and secure multi-party computation algorithms have also been reviewed. The performance analysis of the algorithms has been concentrated in connection with the classification. Classification Rule Mining algorithms are based on centralized data model that is all data is gathered into a single site. Hidenao Abe [4] described a classification rule mining framework by combining the two models i.e., temporal pattern extraction and rule mining. This framework has been developed for mining if-then rules consisting of temporal patterns in left hand side of the rules. The right hand side helps us to predict both of important events and temporal patterns of important index.

Classification algorithms discussed by Hanady Abdulsalam, David B. Skillicorn [5], consist of three phases; a training phase that consists of labeled records, a test phase using previously unseen labeled records, and a deployment phase that classifies unlabeled records. In traditional decision tree classification, a feature (an attribute) of a tuple is either categorical or numerical. Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, and

Sau Dan Lee [6], presented the problem of constructing decision tree classifiers on data with uncertain numerical attributes.

Quinlan J R [7], described a decision trees for classification tasks. These trees are constructed beginning with the root of the tree and proceeding down to its leaves.

Rules can also be extracted from decision trees easily. Many algorithms, such as ID3 and C4.5, have been devised for decision tree construction. These algorithms are widely adopted and used in a wide range of applications as discussed in [6]. Veronica S. Moertini [8] discussed an overview of data classification and its techniques, the basic methods of C4.5 algorithm, the process and the result analysis of the experiment in utilizing C4.5 for varied dataset.

The Decision Tree's can deal with one attribute per test node or with more than one. The former approach is called Univariate Decision Tree, and the second is the Multivariate method. Thales Sehn Korting [9] explains the construction of Univariate DT's and the C4.5 algorithm, used to build such trees. After this, we discuss the Multivariate approach, and how to construct such trees. Based on the analytical evaluation, Salvatore Ruggieri [10] has implemented a more efficient version of the algorithm called C4.5 (Enhanced C4.5 algorithm). It improves on C4.5 by adopting the best among the strategies for computing the information gain of continuous attributes. All the strategies adopt a binary search of the threshold in the whole training set starting from the local threshold computed at a node.

Mahdi Khosravi and Mohammad and J. Tarokh [11] proposed a dynamic mining approach to modeling and predicting users' navigation patterns. Naïve Bayesian algorithm was implemented and shown that this method is effective. This motivate us to present a Naïve Bayesian Classification model for the classification of web log data in quicker time with minimal memory utilization to identify the user preferences more accurately.

## 3. Naive Bayesian Classification Model

The need and requirements of the admin user's of the websites to analyze the user preference become essential, due to massive internet usage. Retrieving the decisive information about the user preferences is achieved, using Naive Bayesian Classification algorithm with quicker time and lesser memory, by means of constructive naïve bayes function. The Naive Bayesian Classification technique as shown in Fig 2, is applied on the web log data to evolve the classification of user page preferences and time spent on the pages of the respective web site (URL).

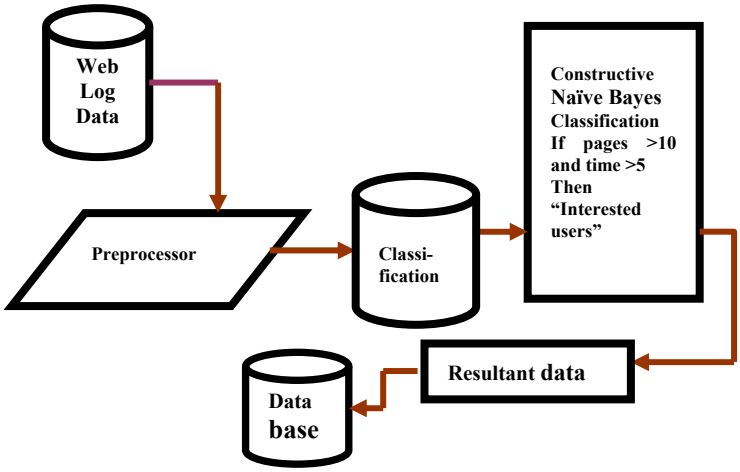


Fig 2 Framework of Naive Bayesian Classification

The web log data (training set) comprises of labels which indicates the class of the observations. New data is classified based on the training set. It preprocesses data in order to remove the irrelevant or redundant attributes and form the normalize data. The decision tree is drawn in a top-down manner. The training sets are at the root. They are partitioned on selected attributes. Partitioning of the training set is processed till there is no more leaf for classifying or there are no samples left and the resultant data is fed to the database.

### 3.1. Data Cleaning Model

All the data obtained from web log are not informative. There may be redundant data or data that are irrelevant for our work. Hence the web log data are fed to the preprocessor which does the work of cleansing to remove the redundant and irrelevant data from the original log file and produces a new file. Then the classification model performs the classification based on the decision tree. Some of the irrelevant information are the images, advertisements and screen savers etc. Based on these data cleaning model the users are identified as "interested users" or "not interested users" using classification by Naive Bayesian Classification model. As a result, data cleaning model has applied the following criteria:

**Failure and aborted requests**: The failure or failed requests are determined using the HTTP code. If the value of the code is 500 the request is said to be a success one, if it does not contain the value as 500 it is derived as a failure request. A page exited due to hardware or software failure is considered as aborted requests.

**No of pages viewed**: Depending on the number of pages viewed by the log file, we can reduce the number of unwanted files. In our work we have considered the minimum pages as 6.

**Time taken**: Depending on the duration of the time spent by the user, we can eliminate the log files which are not used for future references.

### 3.2. Classification Model

Given a training data set, the classification model is used to categorize the given training data set into attributes and the attributes are referred to as class. In our web log data time stamp, users, etc. are considered as attributes or class. Classification can be performed using different techniques. Our work concentrates on decision tree. Our goal is to predict the target class based on our source data (web log data). Our model takes into consideration the binary type of classification in which the target attribute has only two possible variations: for example, interested users or not interested users.

In our work we have used Classification by Naive Bayesian model. It results in less time consumption and less memory utilization. The description about the Naive Bayesian Classification model is described in the forthcoming section.

## 4. Naive Bayesian Classification theorem

Bayesian method is used in decision making that involves probability inferences. This method is more useful when the dimensionality of input is enormous. It uses the prior events to predict the future events. The theorem is explained as given below:

Let Q = $\{x_1,x_2,...,x_n\}$ be a sample training data set whose attributes represent values made on a set of n attributes. Here 'x' is considered as "evidence". Let H represent the hypothesis, in such a way that the data belongs to a specified class C. Our work is to determine P(H | Q). It represents the probability that the hypothesis H holds given the "evidence". For example, our training data set have attributes : session id, time taken, number of pages viewed and that session id is 127.0.0.1 , time taken is 6 minutes and number of pages viewed is 7. Then P(H | Q) is the probability that the session id may be an interested user or not interested user given the time taken and number of pages viewed. Again P(H) is called as priori probability of H. For our example we can represent it as that any session id can be considered as interested or not interested regardless of time taken and number of pages viewed.

4.1. Algorithm

**Initialization**

1. Let T be a training set of samples with k attributes as $A_1,A_2,\ldots..A_k$ given by n dimensional vector $Q = \{x1,x_2,\ldots x_n\}$

2. Let P denotes the probability

3. Let G be the Gaussian distribution value Process

4. Given a sample Q, the classifier performs the prediction to determine the attributes having the highest posteriori probability such that

$$P(A_i \mid Q) > P(A_j \mid Q) \text{ where } i,j = 1,2,\ldots k$$

5. Maximum posteriori hypothesis is calculated using

$$P(A_i \mid Q) = \frac{P(Q \mid A_i)\, P(A_i)}{P(Q)}$$

6. Maximize $P(Q \mid A_i)\, P(A_i)$ if both $P(Q \mid A_i)\, P(A_i)$ are known or $P(Q \mid A_i)$ if only $P(Q \mid A_i)$ is known.

7. If the web log data set contain many attributes it results in maximum of computation time which can be reduced using the following equation

$$P(Q \mid A_i) \equiv \lambda\, P(x_n \mid A_i)$$

8. Calculation of Gaussian distribution with mean μ and standard deviation σ is calculated by

$$G(x, \mu, \sigma) = \frac{1}{\sqrt{3 \prod}} \quad \frac{\exp \quad (x - \mu)^3}{(3\sigma)^3}$$

9. The above equation can be simplified as

$$P(xn \mid A_i) = G(xn, \mu A_i, \sigma A_i)$$

Where $\mu A_i$ refers to the mean and $\sigma A_i$ refers to the standard deviation value of attribute $S_k$.

Maximum Likelihood Estimation (MLE) is used for estimating the parameters for a given training data set. If a clear result cannot be achieved due to time or cost constraint, by using the mean and standard deviation the maximum likelihood estimation can be accomplished. The MLE is discussed in the section given below.

4.2. Maximum likelihood Evaluation (MLE)

Let S be the training set with attributes (s1,s2,…,sn) with a vector as Q. To evaluate the maximum likelihood we have to form the density function that is given as :

$$\mathbf{F(s_1,s_2,\ldots s_n \mid Q) = f(s_1 \mid Q) * f(s_2 \mid Q)*\ldots..f(s_n \mid Q)}$$

$$\mathbf{= \Omega\, f(S_i \mid Q) \text{ where } i = 1,2,\ldots n}$$

Where s1,s2,…sn specifies the parameters and Q being the vector is a random variable. Maximum likelihood for S can be evaluated as

$$\mathbf{Max(\mu A_i, \sigma A_i) \mid \Omega\, f(S_i \mid Q)}$$

In our work Maximum likelihood Evaluation is calculated for a given training data set – web log data which produces a distribution function with the observed data having the greatest probability.

4.3. Decision Tree model

Decision tree model is a method most frequently used in data mining. The purpose is to create a model that predicts the resultant of a target variable based on several input variables given by the user as training data set. An example is shown in fig 3. The interior node corresponds to one of the input variables. The input variable consists of the children as edges. Each leaf shows a value of the target variable given with the values of the input variables which are shown by the path from the root to the leaf. The process is repeated in a recursive manner till a resultant value is derived. The parameters used in web log data to classify the user as interested or not interested is given in table 1.

| Parameters used | Explanation |
|---|---|
| $P_1$ | No of pages view >10 |
| $P_2$ | Time taken >5 |
| $P_3$ | Hyperlink >5 |
| $P_4$ | Personal Information given by user = "yes" |

**Table 1 Parameter consideration**

Using the parameters given in the table 1 a decision tree is formed as in fig 3 using the naïve Bayesian classifier algorithm which helps to determine whether a user who logs into the system is an "interested user" or "not interested user". By means of naïve Bayesian algorithm the memory utilized and time taken can be reduced and maximum likelihood of the parameter is also increased.

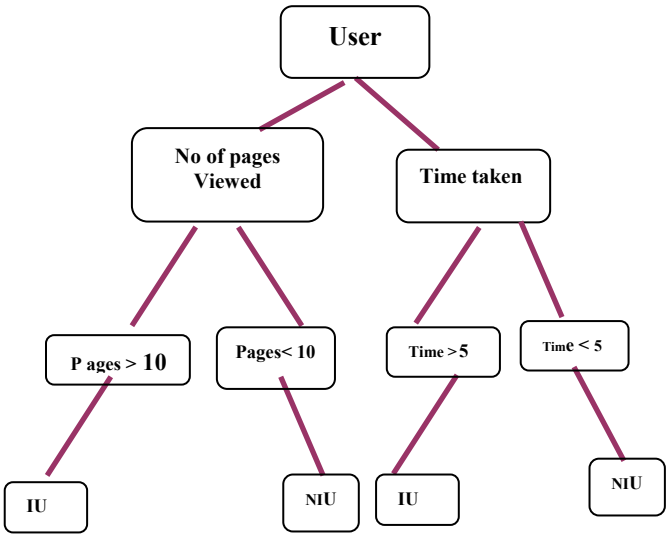



**Fig 3 A Decision Tree generation for Interested User and Not Interested User**

## 5. Experimental Results

Web log data was tested on log files stored by the server. We took into account some portions of log files during the same time interval of five hours of five different days; the sample data set is given in the figure 4.

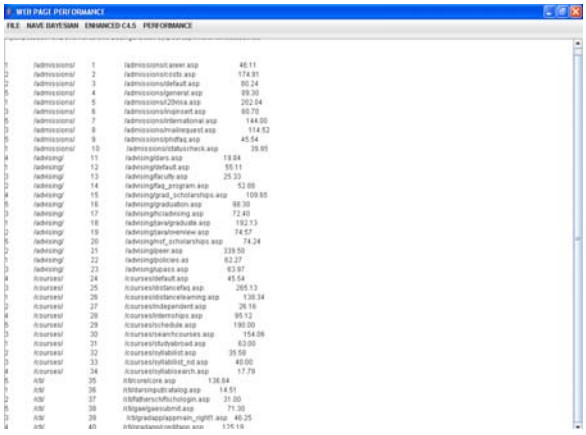


**Fig 4. Sample Log Files**

The data cleaning model evaluate the web log file to determine the log file that are redundant or irrelevant. As described before, the results of data cleaning is derived in 3.1 which removes all group of irrelevant requests.

The C4.5 algorithm is applied on this datasets. The experimental result shows that the time taken by this algorithm is 14.04 Secs and the memory utilization of this algorithm for the same data set is 6.33 KB.

The Naïve Bayesian classification algorithm is applied with the same data sets and experimental results shows that the time taken by this algorithm is 8.68 Secs and memory utilization of this algorithm is 4.35 KB. Which is comparatively efficient than C4.5 algorithm.

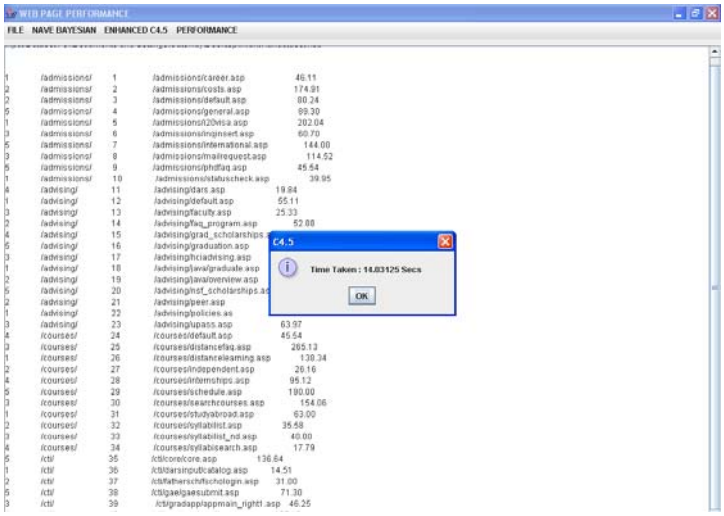



**Fig. 5. Experimental Result that the time taken by C4.5 Algorithm.**

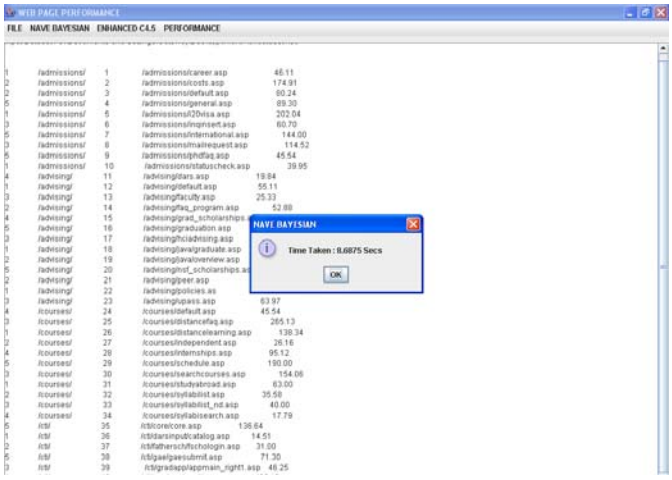



**Fig. 6. Experimental Result that the time taken by Naive Bayesian Classification Algorithm.**

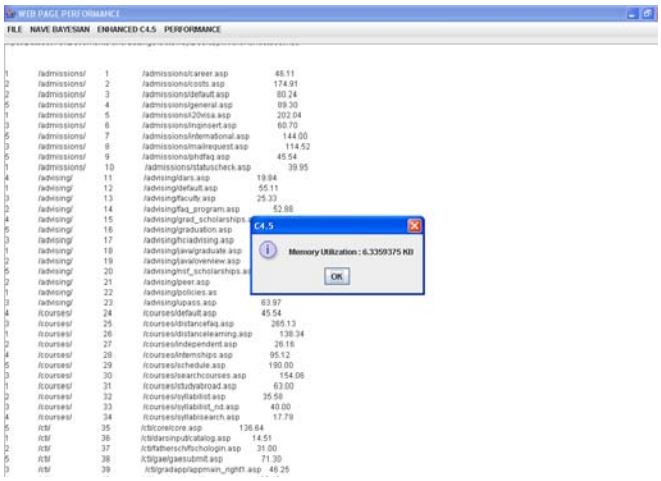


**Fig. 7 Experiment Result for Memory Utilization by C4.5 Algorithm**

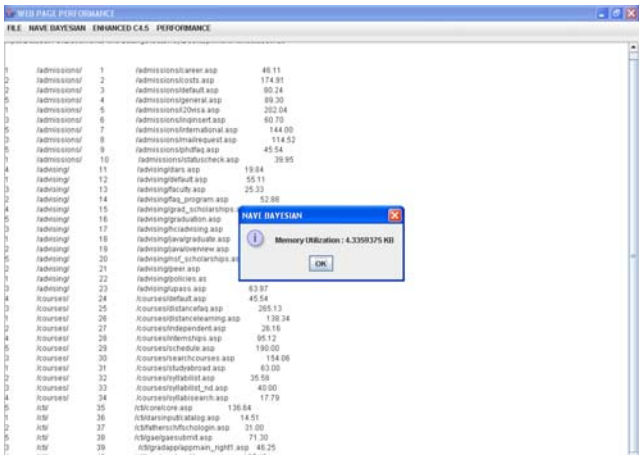


**Fig. 7 Experiment Result for Memory Utilization by Naïve Bayesian Classification Algorithm**

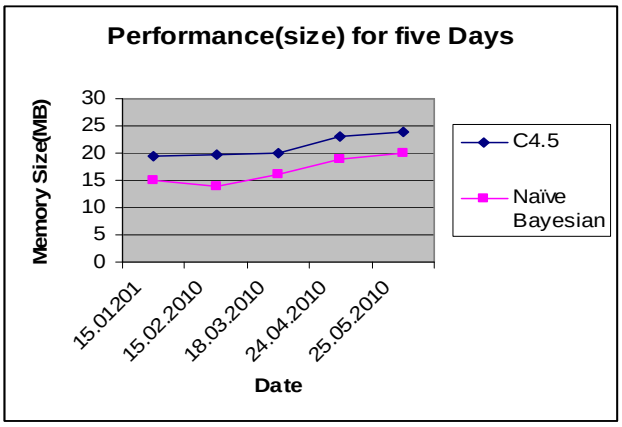



**Fig 9.Comparitive analysis of average performance (size) for C4.5 and Naïve Bayesian Classification**

The figure shows the performance of memory size utilized by the two models. From the graph it is clear that our proposed model Naïve Bayesian Classification outperforms the existing model C4.5 algorithm in terms of minimal memory consumption.

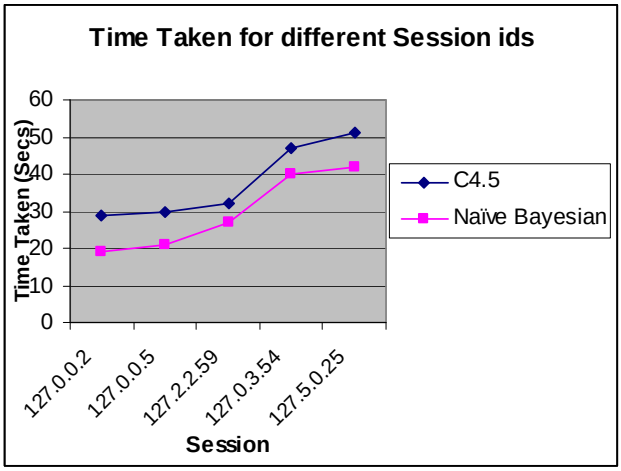



**Fig 10. Comparative study for different session ids**

Figure 10 shows on the x axis the session id and time taken measured on seconds on y axis for web log data file observed on five different days for different session ids. It shows that the average time taken to compute the maximum likelihood of user preference evaluation in Naïve Bayesian Classification for different session ids is better, when compared with the existing model of enhanced C4.5 decision tree algorithm.

## 6. Conclusion

Naive Bayesian Classification shows good result in the improvement in time and memory utilization, it can be applied to any web log files. From the experiments conducted, many attributes are not used for classifying as they are irrelevant. With the help of classification the number of irrelevant attributes can be reduced so that the performance can also be proved efficient. We have given a mathematical evaluation of maximum likelihood for the Naive Bayesian Classification which provided a more efficient implementation with a performance increase compared to enhanced C4.5 decision tree. This method can be used in e-commerce applications, such as Web Caching, Web page recommendation, and Web personalization.

## References

[1] Jie Zhang and Ali., A. Ghorbani, "The Reconstruction of User Sessions from a Server Log Using Improved Time-oriented Heuristics", Proceedings of the Second Annual Conference on Communication Networks and Services, IEEE, May 2004, pp. 315-322.

[2] Mahesh Thylore Ramakrishna1, Latha Kolal Gowdar, Malatesh Somashekar Havanur and Banur Puttappa Mallikarjuna Swamy, "Web Mining: Key Accomplishments, Applications and Future Directions", International Conference on Data Storage and Data Engineering, February 2010.

[3] Alka Gangrade□, Durgesh Kumar Mishra and Ravindra Patel ,"Classification Rule Mining through SMC for Preserving Privacy Data Mining: A Review", International Conference on Machine Learning and Computing IPCSIT, Singapore, 2009, vol.3 (2011), IACSIT Press, pp. 431 to 434.

[4] Hidenao Abe , "Development of a Classification Rule Mining Framework by Using Temporal Pattern Extraction", New fundamental technologies in data mining, January 2011, pp. 493-504.

[5] Hanady Abdulsalam, David B. Skillicorn, and Patrick Martin, "Classification Using Streaming Random Forests",IEEE Transactions on Knowledge And Data Engineering, January 2011, Vol. 23, No.1., pp.22-36.

[6] Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, and Sau Dan Lee, "Decision Trees for Uncertain Data", IEEE Transactions On Knowledge And Data Engineering, January 2011, Vol. 23, No. 1, pp 63-78.

[7] J.R, Quinlan , "Induction of Decision Trees", Machine Learning , 1985, Academic Publishers, pp. 81-106.

[8] Veronica S. Moertini, "Towards The Use Of C4.5 Algorithm For Classifying Banking Dataset", Integral , 2003, Vol. 8 No. 2.

[9] Thales Sehn Korting, "C4.5 algorithm and Multivariate Decision Trees", Image Processing Division, National Institute for Space Research – INPE Sao Jose dos Campos – SP, Brazil, 2006.

[10] Salvatore Ruggieri, "Efficient C4.5", IEEE Transactions on Knowledge and Data Engineering, March/ April 2002, Vol. 14, No. 2, pp. 434 -444.

[11] Mahdi Khosravi, Mohammad and J. Tarokh, "Dynamic Mining of Users Interest Navigation Patterns Using Naive Bayesian Method", IEEE 6th International Conference on Intelligent Computer Communication and Processing Transaction, 2010, pp. 119-122.

**A. K. Santra** received the P. G. degree and Doctorate degree from I.I.T., Kharagpur in the year 1975 and 1981 respectively. He has got 20 years of Teaching Experience and 19 years of Industrial (Research) Experience. His area of interest includes Artificial Intelligence, Neural Networks, Process Modeling, Optimization and Control. He has got to his credit (i) 42 Technical Research Papers which are published in National / International Journals and Seminars of repute, (ii) 20 Research Projects have been completed in varied application areas, (iii) 2 Copy Rights for Software Development have been obtained in the area of Artificial Neural Networks (ANN) and (iv) he is the contributor of the book entitled **"Mathematics and its Applications in Industry and Business"**, Narosa Publishing House, **New Delhi.** He is the recognized Supervisor for guiding Ph. D. / M. S. (By Research) Scholars of Anna University-Chennai, Anna University-Coimbatore, Bharathiyar University, Coimbatore and Mother Teresa University, Kodaikanal. Currently he is guiding 12 Ph. D. Research Scholars in the Department. He is a Life member of CSI and a Life member of ISTE.

**S. Jayasudha** received her M. C. A., from Periyar University, Salem, M.Phil., from Bharathidasan University , Trichy. Currently she is working as Asst.Professor in Bannari Amman Institute of Technology, Sathyamangalam. Her area of interest includes Web Mining, Text Mining. She is a Life member of Computer Society of India and a Life member of Indian Society for Technical Education.

# Design and Development of Artificial Neural Network Based Tamil Unicode Symbols Identification System

**A.B.Karthick Anand Babu**[1]

[1] **Assistant Professor, Department of Software Engineering , Periyar Maniammai University, Vallam,Thanjavur, Tamilnadu,India**

**Abstract**

Design and Development of Unicode and its recognition especially for Indian script is an active area of research today. An attempt is made to identify Tamil- a vernacular of southern India, which is also the official language of Tamilnadu. Tamil language present great challenges to an OCR designer due to the large number (247 letters) in the alphabet, the sophisticated ways in which they combine, and the complicated graphemes they result in. The conventional programming methods of mapping symbol images into matrices, analyzing pixel and/or vector data and trying to decide which symbol corresponds to which character would yield little or no realistic results. Clearly the needed methodology will be one that can detect closeness of graphic representations to known symbols based on the character height, character width, the number of horizontal lines (long and short), the number of vertical lines (long and short), number of slope lines, special dots and based on that the glyphs are now set ready for classification. The extracted features are passed to neural network where the characters are classified by supervised learning of Back Propagation algorithm which compromises training, calculation of error, and modifying weights and then testing the given image.and make decisions based on this nearness. This proposed work has employed the MLP technique to identify the symbols, excellent results were obtained for a number of widely used Unicode Tamil font types.

*Keywords:*
*Artifical Neural Network , MLP , Unicode , Weights*

## 1. Introduction

### 1.1. Artificial Neural Networks

Artificial Neural networks have seen an explosion of interest over the last few years, and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics. Indeed, anywhere that there are problems of prediction, classification or control, neural networks are being introduced. To capture the essence of biological neural systems, an artificial *neuron* is defined as follows:

- It receives a number of inputs (either from original data, or from the output of other neurons in the neural network). Each input comes via a connection that has a strength (or *weight*); these weights correspond to synaptic efficacy in a biological neuron. Each neuron also has a single threshold value. The weighted sum of the inputs is formed, and the threshold subtracted, to compose the *activation* of the neuron.
- The activation signal is passed through an activation function (also known as a transfer function) to produce the output of the neuron.

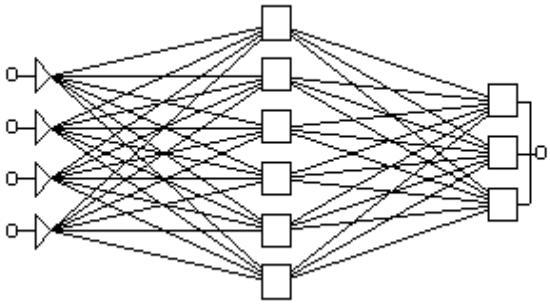


Fig.1 A Typical Feedforward Network

A typical feedforward network has neurons arranged in a distinct layered topology. The input layer is not really neural at all: these units simply serve to introduce the values of the input variables. The hidden and output layer neurons are each connected to all of the units in the preceding layer. Again, it is possible to define networks that are partially-connected to only some units in the preceding layer; however, for most applications fully-connected networks are better.

### 1.2. The Multi-Layer Perceptron Neural Network Model

The Multi-Layer Perceptron Neural Network is perhaps the most popular network architecture in use today. The units each perform a biased weighted sum of their inputs and pass this activation level through an activation function to produce their output, and the units are arranged in a layered feed forward topology. The network thus has a simple interpretation as a form of input-output model, with the weights and thresholds (biases) the free parameters of the model. Such networks can model functions of almost arbitrary complexity, with the number of layers, and the

number of units in each layer, determining the function complexity. Important issues in Multilayer Perceptrons (MLP) design include specification of the number of hidden layers and the number of units in each layer. Most common activation functions are the logistic and hyperbolic tangent sigmoid functions. This work uses the **hyperbolic tangent function:** $f(x) = \frac{2}{(1+e^{-x})} - 1$ (1)

**and derivative:** $f'(x) = f(x)(1 - f(x))$. (2)

### 1.3. Optical Language - Tamil Symbols

The Tamil script is written from left to right is characterized by having its own written symbolic representations, it has twelve vowels, eighteen consonants and one character, the āytam, which is classified in Tamil grammar as being neither a consonant nor a vowel. The script, however, is syllabic and not alphabetic. The complete script, therefore, consists of the thirty-one letters in their independent form, and an additional 216 combinant letters representing a total 247 combinations of a consonant and a vowel, a mute consonant, or a vowel alone. These combinant letters are formed by adding a vowel marker to the consonant. Some vowels require the basic shape of the consonant to be altered in a way that is specific to that vowel. Others are written by adding a vowel-specific suffix to the consonant, yet others a prefix, and finally some vowels require adding both a prefix and a suffix to the consonant. In every case the vowel marker is different from the standalone character for the vowel. Like other South Asian scripts in Unicode, the Tamil encoding was originally derived from the ISCII standard. Both ISCII and Unicode encode Tamil as an abugida. In an abugida, each basic character represents a consonant and default vowel. Consonants with a different vowel or bare consonants are represented by adding a modifier character to a base character. Each codepoint representing a similar phoneme is encoded in the same relative position in each South Asian script block in Unicode, including Tamil. Although Unicode represents Tamil as an abugida all the pure consonants (consonants with no associated vowel) and syllables in Tamil can be represented by combining multiple Unicode code points.

## 2. Technical Overview

### 2.1. Introduction

The operations of the network implementation in this project can be summarized by the following steps:

***Training phase***

- Analyze image for characters
- Convert symbols to pixel matrices
- Retrieve corresponding desired output character and convert to Unicode
- Lineraize matrix and feed to network
- Compute output , Compare output with desired output Unicode value and compute error
- Adjust weights accordingly and repeat process until preset number of iterations

***Testing phase***

- Analyze image for characters
- Convert symbols to pixel matrices
- Compute output
- Display character representation of the Unicode output

Essential components of the implementation are:

- Formation of the network and weight initialization routine
- Pixel analysis of images for symbol detection
- Loading routines for training input images and corresponding desired output characters in special files
- Loading and saving routines for trained network (weight values)
- Character to binary Unicode and vice versa conversion routines
- Error, output and weight calculation routines

### 2.2. Network Formation

The MLP Network implemented for the purpose of this project is composed of 3 layers, one input, one hidden and one output. The input layer constitutes of 150 neurons which receive pixel binary data from a 10x15 symbol pixel matrix. The size of this matrix was decided taking into consideration the average height and width of character image that can be mapped without introducing any significant pixel noise. The hidden layer constitutes of 250 neurons whose number is decided on the basis of optimal results on a trial and error basis.

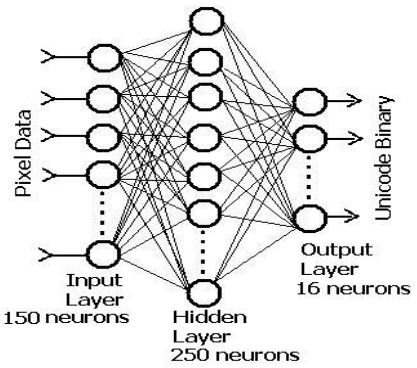



Fig. 2 The Project MLP Network

The output layer is composed of 16 neurons corresponding to the 16-bits of Unicode encoding. To initialize the weights a random function was used to assign an initial random number which lies between two preset integers named **weight_bias**. The weight bias is selected from trial and error observation to correspond to average weights for quick convergence.

## 2.3. Symbol image detection

The Process of Character /symbol Recognition of the document image mainly involves following phases:

- Acquisition and Digitization/Binarization of Grayscale Image
- Thinning and Edge Detection
- Feature Extraction
- Feed Forward Artificial Neural Network based Matching.
- Recognition of Character based on matching score.

A . Detection

The process of image analysis to detect character symbols by examining pixels is the core part of input set preparation in both the training and testing phase. Symbolic extents are recognized out of an input image file based on the color value of individual pixels, which for the limits of this project is assumed to be either black **RGB(255,0,0,0)** or white **RGB(255,255,255,255)**. The input images are assumed to be in bitmap form of any resolution which can be mapped to an internal bitmap. The procedure also assumes the input image is composed of only characters and any other type of bounding object like a border line is not taken into consideration. The procedure for analyzing images to detect characters is listed in the following algorithms:

***i. Determining character lines***

Enumeration of character lines in a character image is essential in delimiting the bounds within which the detection can proceed. Thus detecting the next character in an image does not necessarily involve scanning the whole image all over again.

Algorithm:

1. start at the first x and first y pixel of the image and lines to 0
2. scan up to the width of the image on the same y-component of the image
3. start at the top of the line found and first x-component pixel
4. scan up to the width of the image on the same y-component of the image
5. start below the bottom of the last line found and repeat steps 1-4 to detect subsequent lines
6. If bottom of image (image height) is reached stop.

***ii. Detecting Individual symbols***

Detection of individual symbols involves scanning character lines for orthogonally separable images composed of black pixels.

Algorithm:

1. start at the first character line top and first x-component
2. scan up to image width on the same y-component
3. start at the top of the character found and first x-component, pixel(0,character_top)
4. scan up to the line bottom on the same x-component
5. start at the left of the symbol found and top of the current line, pixel(character_left, line_top)
6. scan up to the width of the image on the same x-component
7. start at the bottom of the current line and left of the symbol, pixel(character_left,line_bottom)
8. scan up to the right of the character on the same y-component


Fig 3. Line and Character boundary detection

From the procedure followed and the above figure it is obvious that the detected character bound might not be the actual bound for the character in question. This is an issue that arises with the height and bottom alignment irregularity that exists with printed unicode symbols. Thus a line top does not necessarily mean top of all characters and a line bottom might not mean bottom of all characters as well. Hence a confirmation of top and bottom for the character is needed. An optional confirmation algorithm implemented in the project is:

A. start at the top of the current line and left of the character
B. scan up to the right of the character
    1. if a black pixels is detected register y as the confirmed top
    2. if not continue to the next pixel
    3. if no black pixels are found increment y and reset x to scan the next horizontal line

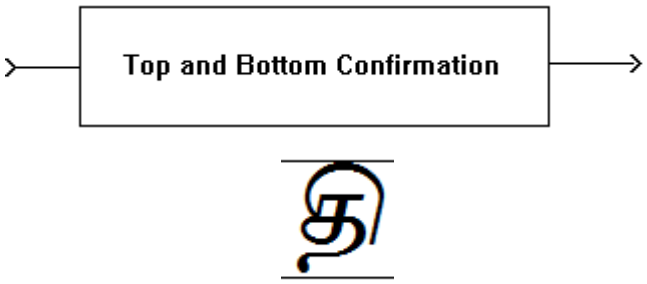



Fig 4. Confirmation of Character boundaries

***iii. Symbol Image Matrix Mapping***

The next step is to map the symbol image into a corresponding two dimensional binary matrix. An important issue to consider here will be deciding the size of the matrix. If all the pixels of the symbol are mapped into the matrix, one would definitely be able to acquire all the distinguishing pixel features of the symbol and minimize overlap with other symbols. However this strategy would imply maintaining and processing a very large matrix (up to 1500 elements for a 100x150 pixel image). Hence a reasonable tradeoff is needed in order to minimize processing time which will not significantly affect the separability of the patterns. The project employed a sampling strategy which would map the symbol image into a 10x15 binary matrix with only 150 elements. Since the height and width of individual images vary, an adaptive sampling algorithm was implemented.

B. Training

Once the network has been initialized and the training input space prepared the network is ready to be trained. Some issues that need to be addressed upon training the network are:

- How chaotic is the input space? A chaotic input varies randomly and in extreme range without any predictable flow among its members.
- How complex are the patterns for which we train the network? Complex patterns are usually characterized by feature overlap and high data size.
- What should be used for the values of:
  - Learning rate
  - Sigmoid slope
  - Weight bias
- How many Iterations are needed to train the network for a given number of input sets?
- What error threshold value must be used to compare against in order to prematurely stop iterations if the need arises?

The complexity of the individual pattern data is also another issue in character recognition. Each symbol has a large number of distinct features that need to be accounted for in order to correctly recognize it. Elimination of some features might result in pattern overlap and the minimum amount of data required makes it one of the most complex classes of input space in pattern recognition. Other than the known issues mentioned, the other numeric parameters of the network are determined in real time. They also vary greatly from one implementation to another according to the number of input symbols fed and the network topology.

For the purpose of this project the parameters use are:

- Learning rate = 170
- Sigmoid Slope = 0.017
- Weight bias = 25 (determined by trial and error)
- Number of Epochs = 200-700 (depends on font )
- Mean error threshold value = 0.0003 (determined by trial and error)

Algorithm:

The training routine implemented the following basic algorithm

1. Form network according to the specified topology parameters
2. Initialize weights with random values within the specified weight_bias value
3. load trainer set files (both input image and desired output text)
4. analyze input image and map all detected symbols into linear arrays
5. read desired output text from file and convert each character to a binary Unicode value to store separately
6. for each character :
   a. calculate the output of the feed forward network
   b. compare with the desired output corresponding to the symbol and compute error
   c. back propagate error across each link to adjust the weights
7. move to the next character and repeat step 6 until all characters are visited
8. compute the average error of all characters
9. repeat steps 6 and 8 until the specified number of epochs
   a. Is error threshold reached? If so abort iteration
   b. If not continue iteration

C. Testing

The testing phase of the implementation is simple and straightforward. Since the program is coded into modular parts the same routines that were used to load, analyze and compute network parameters of input vectors in the training phase can be reused in the testing phase as well. The basic steps in testing input images for characters can be summarized as follows:

Algorithm:

- load image file
- analyze image for character lines
- for each character line detect consecutive character symbols
  - analyze and process symbol image to map into an input vector
  - feed input vector to network and compute output
  - convert the Unicode binary output to the corresponding character and render to a text box

## 3 Results and Discussion

The network has been trained and tested for a number of widely used font type in Tamil Font style. The necessary steps are preparing the sequence of input symbol images in a single image file (*.bmp [bitmap] extension), typing the corresponding characters in a text file (*.cts [character trainer set] extension) and saving the two in the same folder (both must have the same file name except for their extensions). The application will provide a file opener dialog for the user to locate the *.cts text file and will load the corresponding image file by itself. Although the results listed in the subsequent tables are from a training/testing process of symbol images created with a 72pt. font size the use of any other size is also straight forward by preparing the input/desired output set as explained. The application can be operated with symbol images as small as 8pt font size.

*A. Results for variation in number of Epochs*
Number of characters=90, Learning rate=150, Sigmoid slope=0.014

| **Font Type** | **300** | | **600** | | **800** | |
|---|---|---|---|---|---|---|
| | № of wrong characters | % Error | № of wrong characters | % Error | № of wrong characters | % Error |
| Vijaya | 4 | 4.44 | 3 | 3.33 | 1 | 1.11 |
| Latha | 1 | 1.11 | 0 | 0 | 0 | 0 |

Table 1

*B. Results for variation in number of Input characters*
Number of Epochs=100, Learning rate=150, Sigmoid slope=0.014

| **Font Type** | **20** | | **50** | | **90** | |
|---|---|---|---|---|---|---|
| | № of wrong characters | % Error | № of wrong characters | % Error | № of wrong characters | % Error |
| Vijaya | 0 | 0 | 6 | 12 | 11 | 12.22 |
| Latha | 0 | 0 | 3 | 6 | 8 | 8.89 |

Table 2

*C. Results for variation in Learning rate parameter*
Number of characters=90, Number of Epochs=600, Sigmoid slope=0.014

| **Font Type** | **50** | | **100** | | **120** | |
|---|---|---|---|---|---|---|
| | № of wrong characters | % Error | № of wrong characters | % Error | № of wrong characters | % Error |
| Vijaya | 82 | 91.11 | 18 | 20 | 3 | 3.33 |
| Latha | 56 | 62.22 | 11 | 12.22 | 1 | 1.11 |

Table 3

## 4. Performance Observation

### 1. Influence of parameter variation

i. Increasing the number of iterations has generally a positive proportionality relation to the performance of the network. However in certain cases further increasing the number of epochs has an adverse effect of introducing more number of wrong recognitions. This partially can be attributed to the high value of learning rate parameter as the network approaches its optimal limits and further weight updates result in bypassing the optimal state. With further iterations the network will try to swing back to the desired state and back again continuously, with a good chance of missing the optimal state at the final epoch. This phenomenon is known as over learning.

ii. The size of the input states is also another direct factor influencing the performance. It is natural that the more number of input symbol set the network is required to be trained for the more it is susceptible for error. Usually the complex and large sized input sets require a large topology network with more number of iterations. For the above maximum set number of 90 symbols the optimal topology reached was one hidden layer of 250 neurons.

iii. Learning rate parameter variation also affects the network performance for a given limit of iterations. The less the value of this parameter, the lower the value with which the network updates its weights. This intuitively implies that it will be less likely to face the over learning difficulty discussed above since it will be updating its links slowly and in a more refined manner. But unfortunately this would also imply more number of iterations is required to reach its optimal state. Thus a trade of is needed in order to optimize the overall network performance. The optimal value decided upon for the learning parameter is 150.

## 5. Reference

1. **Artificial Intelligence and cognitive science** 2006, Nils J. Nilsson Stanford AI Lab
2. **Off-line Handwriting Recognition Using Artificial Neural Networks** 2000, Andrew T. Wilson University of Minnesota, Morris
3. **Using Neural Networks to Create an Adaptive Character Recognition System** 2002, Alexander J. Faaborg Cornell University, Ithaca NY

4. **Hand-Printed Character Recognizer using Neural Network**
2000, Shahzad Malik
5. **Neural Networks and Fuzzy Logic**. 1995, Rao, V., Rao, H.MIS Press, New York
6. **"Character Recognition by Neural Network,"** in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 196–201 G. Guodong, S. Li, and C. Kapluk
7. **"A Devnagari OCR and A Brief Overview of OCR for Indian Script"**, *PROC Symposium on Transaction support System (STRANS 2001)*, Feb. 15-17, 2001, Kanpur, India Veena Bansal and R.M.K. Sinha, "A Devnagari OCR and A Brief Overview of OCR for Indian Script", *PROC Symposium on Transaction support System (STRANS 2001)*,Feb. 15-17, 2001, Kanpur, India.
8. http://en.wikipedia.org/wiki/Tamil_alphabet

9. ABOUT THE AUTHOR

10.

A.B.Karthick Anand Babu working as Assistant Professor in the department of Software Engineerin, Periyar Maniammai University,Vallam,Thanjavur. He is currently working in the area of effective teaching and easy learning methodology.

# Hybrid DCT-DWT Watermarking and IDEA Encryption of Internet Contents

**M.A. Mohamed and A.M. El-Mohandes**

**Electronics and Communication Engineering, Faculty of Engineering-Mansoura University
Mansoura, Dakhlia, Egypt**

**Abstract**

Encryption and watermarking are complementary lines of defense in protecting multimedia content. Recent watermarking techniques have therefore been developed independent from encryption techniques. In this paper, we present a hybrid image protection scheme to establish a relation between the data encryption key and the watermark. Prepositioned secret sharing allows the reconstruction of different encryption keys by communicating different activating shares for the same prepositioned information. Each activating share is used by the receivers to generate a fresh content decryption key. In the proposed scheme, the activating share is used to carry copyright or usage rights data. The bit stream that represents this data is also embedded in the content as a visual watermark. When the encryption key needs to change, the data source generates a new activating share, and encrypts the corresponding data with the key constructed from the new activating share. Before transmission, the encrypted data is embedded in a multimedia stream. Each receiver can extract the encrypted data from the host image, and decrypt this data after reconstructing the same key. Our presentation will include the application of the scheme to a test image, and a discussion on the data hiding capacity, watermark transparency, and robustness to common attacks.

***Keywords:*** *discrete cosine transform, discrete wavelet transform, and international data encryption algorithm (IDEA), Bit correct ratio.*

## 1. Introduction

The advent of the Internet has resulted in many new opportunities for the creation and delivery of content in digital form. Applications include electronic advertising, real-time video and audio delivery, digital repositories and libraries, and Web publishing. An important issue that arises in these applications is the protection of the rights of all participants [1], [2]. It has been recognized for quite some time that current copyright laws are inadequate for dealing with digital data. This has led to an interest towards developing new copy deterrence and protection mechanisms. One such effort that has been attracting increasing interest is based on digital watermarking techniques [3]. Digital watermarking; Fig.1, is the process of embedding information into digital multimedia content such that the information (which we call the watermark) can later be extracted or detected for a variety of purposes including copy prevention and control. Digital watermarking has become an active and important area of research, and development and commercialization of watermarking techniques is being deemed essential to help address some of the challenges faced by the rapid proliferation of digital content [4]. A secure computing environment would not be complete without consideration of encryption technology; Fig.2. The term encryption refers to the practice of obscuring the meaning of a piece of information by encoding it in such a way that it can only be decoded, read and understood by people for whom the information is intended. It is the process of encoding data to prevent unauthorized parties from viewing or modifying it.

The use of simple codes to protect information can be traced back to the fifth century BC. As time has progressed, the methods by which information is protected have become more complex and more secure. Encryption can be used to provide high levels of security to network communication, e-mail, files stored on hard drives or floppy disks, and other information that requires protection. Encryption and watermarking each provide a different line of defense in protecting content. Recent research has therefore followed two different avenues resulting in encryption techniques that are independent from watermarking techniques: (i) Encryption makes the content unintelligible through a reversible mathematical transformation based on a secret key [1], [2]. In secure multimedia content distribution, the audio/visual stream is compressed, packetized and encrypted [3].

One of the most challenging problems in distribution architectures is the delivery of the decryption key, and (ii) Watermarking (data hiding) [4] – [7] is the process of embedding data into a multimedia element such as image, audio or video. This embedded data can later be extracted from, or detected in, the multimedia element for different purposes such as copyright protection, access control, and broadcast monitoring. One possible avenue of research is to establish a relation between the data encryption key and the watermark. A recent method for delivering the keying information in secure multimedia multicast applications suggests the use of a media dependent channel [2]. The rekey messages are embedded in the multimedia stream rather than being sent in a separate channel.

This paper is organized as the following: section (2) introduces the performance measures of: digital watermarking techniques as well as the combined digital watermarking and encryption techniques; (3) discusses the DCT-DWT as a combined digital watermarking technique; (4) presents the basics of encryption and the state of the art of international data encryption algorithm (IDEA) technique; (5) provides the simulation results, and (6) introduces the final conclusion of this paper.

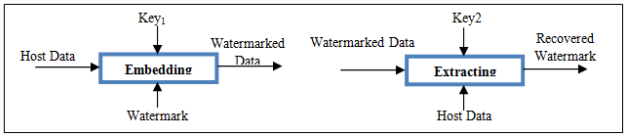



Fig.1 Digital watermarking

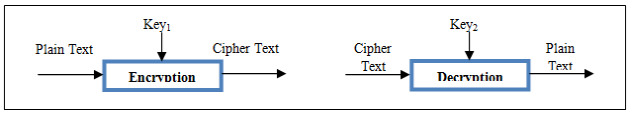



Fig.2 Digital encryption

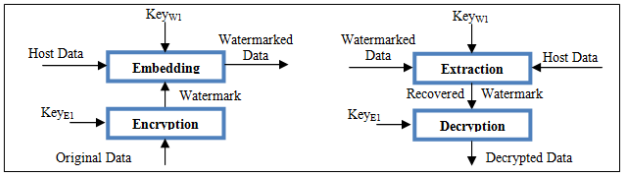



Fig.3 Hybrid technique

## 2. Performance Measures

How then do we provide metrics for the evaluation of the embedding techniques? Capacity and speed can be easily evaluated using the number of bits per cover size, and computational complexity, respectively. The systems use of keys is more or less by definition, and the statistical imperceptibility by correlation between the original images and recovered counterpart. The more difficult task is providing metrics for perceptibility and robustness.

### 2.1 Perceptibility Measures

There are five measures used in evaluating the imperceptibility of any embedding system. Assume X is the cover image, and $X_W$ is the watermarked image. Whereas, W is the watermark, and W' is the recovered watermark each of dimensions M×N. It should be noted that these performance measures could be evaluated with respect to: the watermarked data and the host data.

#### 2.1.1 Capacity

$$C = B_w / B_i \tag{1}$$

where $B_w$ is number of watermark bits and Bi is number of host bits.

#### 2.1.2 Minimum Square Error (MSE)

$$MSE = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(W(i,j) - W'(i,j)\right)^2 / (M \cdot N) \tag{2}$$

#### 2.1.3 Peak Signal-to-Noise Ratio (PSNR)

$$PSNR = 20 \times Log\left[\left(2^n - 1\right)/ v\right] \tag{3}$$

where n is the number of bits per pixel.

#### 2.1.4 Correlation Coefficients (R)

$$R = \sum_{i=1}^{N}\left(Xi - \overline{X}\right)\sum_{i=1}^{N}\left(Yi - \overline{Y}\right)\Big/\left(N \bullet \sum_{i=1}^{N}\left(Xi - \overline{X}\right)^2 . \sum_{i=1}^{N}\left(Yi - \overline{Y}\right)^2\right) \tag{4}$$

where N is the number of pixels, Xi is the pixel value of the original image, $\overline{X}$ is the average value of the original image; Y is the pixel value of the modified image and $\overline{Y}$ is the average value of the modified image.

#### 2.1.5 Watermark-to-Document Ratio (WDR)

$$WDR = 10 \times Log\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X(i,j) - X'(i,j)\right)^2 / \sum_{i=1}^{M}\sum_{j=1}^{N}X^2(i,j)\right) \tag{5}$$

### 2.2 Measuring Robustness

Low level is the bare minimum requirements that a watermark must meet in order to be considered useful. Watermarks at this level should be resistant to common modifications that non-malicious users with inexpensive tools might do to images. As the robustness increases more specialized and expensive tools become required, as well as more intimate knowledge of the watermarking system being used. At the very top of the scale is provable reliability in which it is either computationally or mathematically impossible to remove or disable the mark [8]. Some of the early literature considered a binary robustness metric that only allows for two different states. However, it makes sense to use a metric that allows for different levels of robustness. The use of bit-correct ratio (BCR) has become recently, as it allows for more detailed scale of values; BCR is defined as:

$$BCR = \frac{100}{l}\sum_{n=0}^{l-1}\begin{cases}1 & W_n' = W_n \\ 0 & W_n' \neq W\end{cases} \tag{6}$$

where l is the watermark length, $W_n$ is the nth bit of the embedded watermark and $W_n$' is the nth bit of the recovered watermark. In this paper, some data processing operations are applied to the watermarked data; these are: (i) JPEG compression; (ii) mean filter; (iii) median filter;

(iv) Gaussian noise; (v) blurring, (vi) histogram equalization; (vii) Gamma correction, and (viii) intensity adjustment.

## 3. Combined Technique

Commonly used frequency-domain transforms include the Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT). However, DWT [7] has been used in digital image watermarking more frequently due to its excellent spatial localization and multi-resolution characteristics, which are similar to the theoretical models of the human visual system [8]. Further performance improvements in DWT-based digital image watermarking algorithms could be obtained by combining DWT with DCT [7]. The idea of applying two transforms is based on the fact that combined transforms could compensate for the drawbacks of each other, resulting in effective watermarking.

### 3.1 The DCT and DWT Transforms

The discrete cosines transform (DCT) and discrete wavelet transform (DWT) transforms have been extensively used in many digital signal processing applications. In this section, we introduce the two transforms briefly, and outline their relevance to the implementation of digital watermarking. The DCT is a technique for converting a signal into elementary frequency components [9]. It represents an image as a sum of sinusoids of varying magnitudes and frequencies. With an input image, x, the DCT coefficients for the transformed output image, y, are computed according to Eq.7 shown below. In the equation, x, is the input image having N x M pixels, x(m, n) is the intensity of the pixel in row m and column n of the image, and y(u, v) is the DCT coefficient in row u and column v of the DCT matrix.

$$y(u,v)=\sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}}\alpha_u\alpha_v\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(x(m,n)\cos(\frac{(2m+1)u\pi}{2M})\cos(\frac{(2n+1)v\pi}{2N})) \quad (7)$$

where

$$\alpha_u = \begin{cases} 1/\sqrt{2} & u=0 \\ 1 & u=1,2,.....,\ M-1 \end{cases}$$

$$\alpha_v = \begin{cases} 1/\sqrt{2} & v=0 \\ 1 & v=1,2,.....,\ N-1 \end{cases} \quad (8)$$

The image is reconstructed by applying inverse DCT operation according to Eq.9:

$$x(m,n)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}(\sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}}\alpha_u\alpha_v y(u,v)\cos(\frac{(2m+1)u\pi}{2M})\cos(\frac{(2n+1)v\pi}{2N})) \quad (9)$$

The popular block-based DCT transform segments image non-overlapping blocks and applies DCT to each block. These results in giving three frequency sub-bands: low frequency sub-band, mid-frequency sub-band and high frequency sub-band. DCT-based watermarking is based on two facts. The first fact is that much of the signal energy lies at low-frequencies sub-band which contains the most important visual parts of the image. The second fact is that high frequency components of the image are usually removed through compression and noise attacks. The watermark is therefore embedded by modifying the coefficients of the middle frequency sub-band so that the visibility of the image will not be affected and the watermark will not be removed by compression.

The DWT transform: Wavelets are special functions which, in a form analogous to sines and cosines in Fourier analysis, are used as basal functions for representing signals [10]. For 2D images, applying DWT corresponds to processing the image by 2D filters in each dimension. The filters divide the input image into four non-overlapping multi-resolution sub-bands LL1, LH1, HL1 and HH1. The sub-band LL1 represents the coarse-scale DWT coefficients while the sub-bands LH1, HL1 and HH1 represent the fine-scale of DWT coefficients. To obtain the next coarser scale of wavelet coefficients, the sub-band LL1 is further processed until some final scale N is reached. When N is reached we will have 3N+1 sub-bands consisting of the multi-resolution sub-bands LLN and LHx, HLx and HHx where x ranges from 1 until N. Due to its excellent patio-frequency localization properties, the DWT is very suitable to identify the areas in the host image where a watermark can be embedded effectively. In particular, this property allows the exploitation of the masking effect of the human visual system such that if a DWT coefficient is modified, only the region corresponding to that coefficient will be modified.

In general most of the image energy is concentrated at the lower frequency sub-bands LLx and therefore embedding watermarks in these sub-bands may degrade the image significantly. Embedding in the low frequency sub-bands, however, could increase robustness significantly. On the other hand, the high frequency sub-bands HHx include the edges and textures of the image and the human eye is not generally sensitive to changes in such sub-bands. This allows the watermark to be embedded without being perceived by the human eye. The compromise adopted by many DWT-based watermarking algorithm, is to embed the watermark in the middle frequency sub-bands LHx and HLx where acceptable performance of imperceptibility and robustness could be achieved.

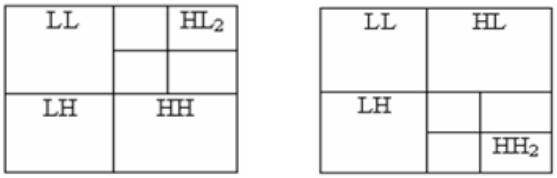



Fig.4 DWT decomposition levels

### 3.2 The Combined DCT-DWT Algorithm

The watermark embedding procedure is depicted in Fig.5 followed by a detailed explanation [11-14]:

**Step-1:** Apply DWT to decompose the cover host image into four non-overlapping multi-resolution sub-bands: LL1, HL1, LH1, and HH1.

**Step-2:** Apply DWT again to sub-band HL1 to get four smaller sub-bands and choose the HL2 sub-band as shown in Fig.4a. Or, apply DWT to sub-band HH1 to get four smaller sub-bands and choose the HH2 sub-band as shown in Fig.4b.

**Step-3:** Divide the sub-band HL2 (or HH2) into 8 x 8 blocks.

**Step-4:** Apply DCT to each block in the chosen sub-band (HL2 or HH2).

**Step-5:** Re-formulate the grey-scale watermark image into a vector of zeros and ones.

**Step-6:** Generate two uncorrelated pseudorandom sequences. One sequence is used to embed the watermark bit 0 (PN_0) and the other sequence is sued to embed the watermark bit 1 (PN_1). Number of elements in each of the two pseudorandom sequences must be equal to the number of mid-band elements of the DCT-transformed DWT sub-bands.

**Step-7:** Embed the two pseudorandom sequences, PN_0 and PN_1, with a gain factor α, in the DCT transformed 8x8 blocks of the selected DWT sub-bands of the host image. Embedding is not applied to all coefficients of the DCT block, but only to the mid-band DCT coefficients. If we donate $\boldsymbol{X}$ as the matrix of the mid-band coefficients of the DCT transformed block, then embedding is done as follows: If the watermark bit is 0 then:

$$X' = X + (\alpha \times PN\_0) \quad (10)$$

Otherwise, if the watermark bit is 1 then,

$$X' = X + (\alpha \times PN\_1) \quad (11)$$

**Step-8:** Apply inverse DCT (IDCT) to each block after its mid-band coefficients have been modified to embed the watermark bits as described in the previous step.

**Step 9:** Apply the inverse DWT (IDWT) on the DWT transformed image, including the modified sub-band, to produce the watermarked host image.

The watermark *extraction* procedure is depicted in Fig.6, and described in details in the following steps. The combined DWT-DCT algorithm is a blind watermarking algorithm, and thus the original host image is not required to extract the watermark:

**Step-1:** Apply DWT to decompose the watermarked image into four non-overlapping multi-resolution subbands: LL1, HL1, LH1, and HH1.

**Step-2:** Apply DWT to HL1 to get four smaller subbands, and choose the sub-band HL2, as shown in Fig. 4a. Or, apply DWT to the HH1 sub-band to get four smaller sub-bands, and choose the HH2 sub-band, as shown in Fig. 4b.

**Step-3:** Divide the sub-band HL2 (or HH2) into 8 x 8 blocks.

**Step-4:** Apply DCT to each block in the chosen sub-band (HL2 or HH2), and extract the mid-band coefficients of each DCT transformed block.

**Step-5:** Regenerate the two pseudorandom sequences (PN_0 and PN_1) using the same seed used in the watermark embedding procedure.

**Step-6:** For each block in the sub-band HL2 (or HH2), calculate the correlation between the mid-band coefficients and the two generated pseudorandom sequences (PN_0 and PN_1). If the correlation with the PN_0 was higher than the correlation with PN_1, then the extracted watermark bit is considered 0, otherwise the extracted watermark is considered 1.

**Step-7:** Reconstruct the watermark using the extracted watermark bits, and compute the similarity between the original and extracted watermarks.

## 4. Encryption

Encryption is said to occur when data is passed through a series of mathematical operations that generate an alternate form of that data; the sequence of these operations is called an algorithm. To help distinguish between the two forms of data, the unencrypted data is referred to as the plaintext and the encrypted data as ciphertext [1, 15]. The security of encryption lies in the ability of an algorithm to generate ciphertext that is not easily reverted to the original plaintext. Encryption is generally used because of its benefits such as ease –of– use, reliability, and security. Encryption divided into many categories: (i) according to key; we have two types (Symmetric key and Asymmetric key algorithm); (ii) according to the way of encryption; we have two types (Stream ciphering and Block ciphering), and (iii) according to the type of operation; we have two types (Substitution and Transposition) [2].

### 4.1 Why we use Encryption?

There are many reasons of using encryption: (i) Reliability means that the cipher text will always be recoverable and the recovered data will be the same as to the original plaintext; (ii) Security means that the encryption system will in fact keep the information hidden from all but those persons intended to see it, and (iii) An encryption system which is easy to use should allow keyboard entry of a string from 10 to 60 characters Ease-of-use is easy to understand.

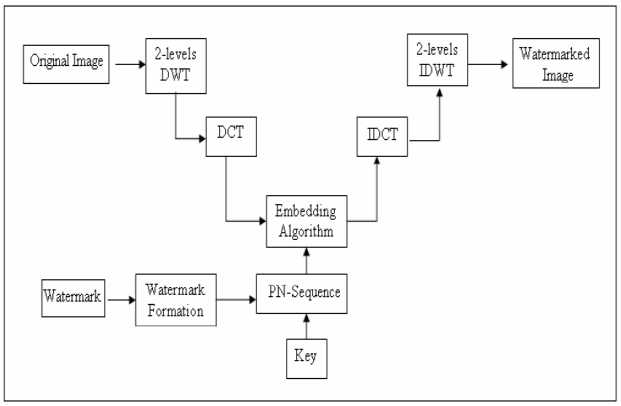



Fig.5 Embedding in DCT-DWT technique

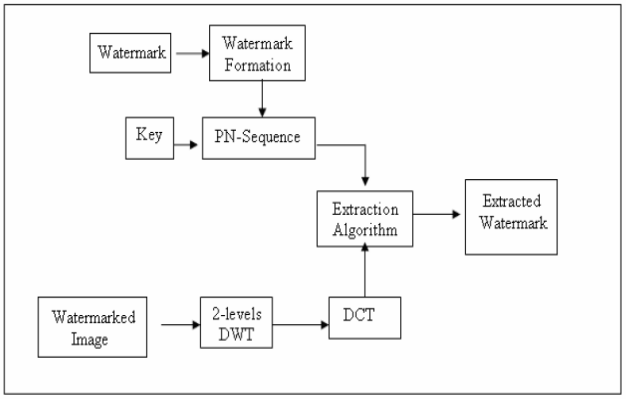



Fig.6 Extraction in DCT-DWT technique

## 4.2 Advantages & Disadvantages of Encryption

Encryption can play a very important role in your day-to-day computing and communicating: (i) Encryption can protect information stored on your computer from unauthorized access - even from people who otherwise have access to your computer system; (ii) Encryption can protect information while it is in transit from one computer system to another, and (iii) Encryption can be used to deter and detect accidental or intentional alterations in your data. Despite these advantages, encryption has its limits: (i) Encryption can't prevent an attacker from deleting your data altogether; (ii) An attacker might find a previously unknown and relatively easy way to decode messages encrypted with the algorithm you are using, and (iii) An attacker could access your file before it is encrypted or after it is decrypted [15-20].

## 4.3 Encryption Benefits

In addition to the advantages of encryption; it has the following benefits: (i) data confidentiality-protects the privacy and safety of business and personal information; (ii) Access control–allows data access only to authorized users, and (iii) key management–provides key generation, distribution, and storage.

## 4.4 IDEA Algorithm

International data encryption algorithm (IDEA) is patented by the Swiss firm of Ascom. They have, however, been generous in allowing, with permission, free noncommercial use of their algorithm, with the result that IDEA is best known as the block cipher algorithm used within the popular encryption program PGP. The IDEA algorithm is interesting in its own right. It includes some steps which, at first, make it appear that it might be a non-invertible hash function instead of a block cipher [1], [2]. Also, it is interesting in that it entirely avoids the use of any lookup tables or S-boxes. IDEA uses 52 subkeys, each 16 bits long. Two are used during each round proper, and four are used before every round and after the last round. It has eight rounds; Fig.7 and 8.

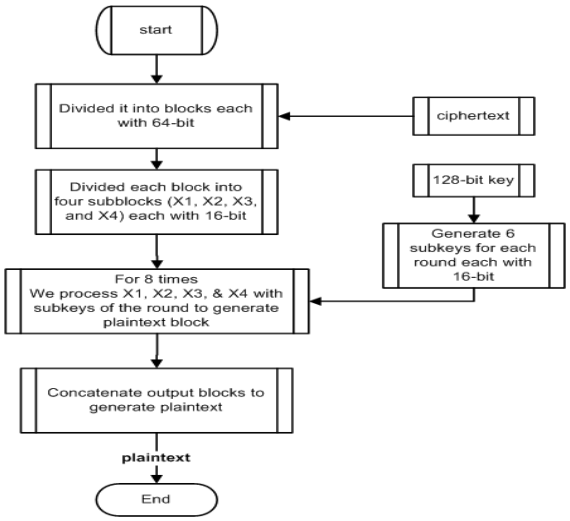



Fig.7The encryption process of IDEA

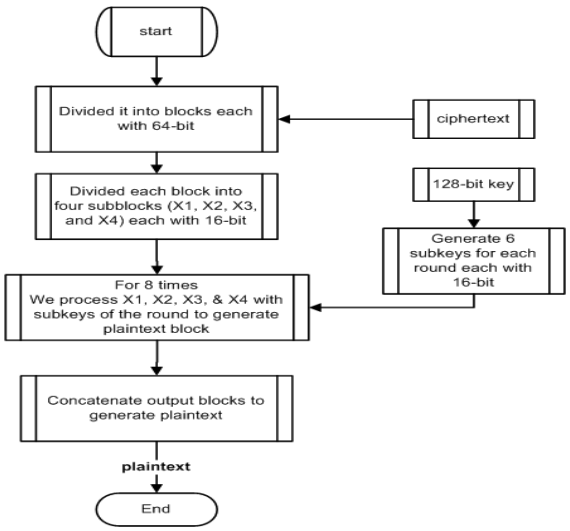



Fig.8The decryption process of IDEA

The plaintext block in IDEA is divided into four quarters, each 16 bits long. Three operations are used in IDEA to combine two 16 bit values to produce a 16 bit result, addition, XOR, and multiplication. Addition is normal addition with carries, modulo 65,536. Multiplication, as used in IDEA, requires some explanation. Multiplication by zero always produces zero, and is not invertible. Multiplication modulo n is also not invertible whenever it is by a number which is not relatively prime to n. The way multiplication is used in IDEA, it is necessary that it be always invertible. This is true of multiplication IDEA style. The number 65,537, which is $2^{16}+1$, is a prime number. (Incidentally, $2^8+1$, or 257, is also prime, and so is $2^4+1$, or 17, but $2^{32}+1$ is not prime, so IDEA cannot be trivially scaled up to a 128-bit block size.) Thus, if one forms a multiplication table for the numbers from 1 through 65,536, each row and column will contain every number once only, forming a Latin square, and providing an invertible operation. The numbers that 16 bits normally represent are from 0 to 65,535 (or, perhaps even more commonly, from -32,768 to 32,767). In IDEA, for purposes of multiplication, a 16 bit word containing all zeroes is considered to represent the number 65,536; other numbers are represented in conventional unsigned notation, and multiplication is modulo the prime number 65,537.

## 5. Results and Discussions

The performance of two combined digital watermarking techniques: (i) DCT-DWT and (ii) modified DCT-DWT; with and without encryption was tested and evaluated. The simulation process was performed using MATLAB-2009a using Intel Core2 Duo 2.66 GHz, 4 GB RAM on Microsoft Windows Service Pack-3. The host image was chosen as a grayscale image of size 256×256, uncompressed, and of 8 bits per pixel depth. The watermark data was simulated as a black and white image of size 32×128 containing text information. The results provided in Tables 1 introduce the objective evaluation parameters: (i) capacity; (ii) MSE; (iii) $PSNR_{dB}$; (iv) $WDR_{dB}$, and (v) R.

These results show a great improvement in the performance of the combined watermarking technique when assigned with encryption. In addition the use of hybrid system of digital watermarking and encryption resist different types of attacks in a very good manner as shown in the BCR values presented in Table 2. These results showed that the proposed technique provided a great improvement in resisting: (i) histogram equalization; (ii) gamma correction; (iii) intensity adjustment; (iv) blurring; (v) JPEG compression; (vi) cropping; (vii) median filtering, and (viii) mean filtering. These improvements are illustrated in the set of figures from Fig.9 to Fig.22. Fig.9 shows the watermark information; Fig.10 shows the host image; Fig.11 shows the watermarked image without encryption; Fig.12 illustrates the recovered watermark; Fig.13 provides the encrypted watermark; Fig.14 illustrates the watermarked data with encryption; Fig.15 shows the recovered encrypted watermark from Fig.14; Fig.16 gives the recovered decrypted watermark of Fig.15; Fig.17 watermarked data with modified technique without encryption; Fig.18 recovered watermark from Fig.17; Fig.19 encrypted original watermark; Fig.20 shows the watermarked data using modified watermarking technique with encryption; Fig.21 recovered encrypted watermark from Fig.20, and Fig.22 shows the decrypted watermark of Fig.21.

## 6. Conclusion

From our results we can easily see that our modified watermarking with encryption technique improves the imperceptibility measures (PSNR, WDR, R, Capacity, and MSE), and also improves the robustness against attacks that depend on filtering the data. Although we can see that our modified technique makes the robustness against JPEG compression attacks, histogram equalization attacks, etc. worst than watermarking without encryption.

Table.1 Results of imperceptibility measures

| Parameter | Watermark Only | | Watermark with Encryption | |
|---|---|---|---|---|
| | DCT-DWT | Modified DCT-DWT | DCT-DWT | Modified DCT-DWT |
| Capacity | 0.0037 | 0.0037 | 0.0037 | 0.0037 |
| MSE | 228.0360 | 2.2903 | 157.1601 | 1.5755 |
| PSNR(dB) | 24.6707 | 45.5215 | 26.2804 | 46.5283 |
| WDR(dB) | -24.5934 | -38.1551 | -26.2123 | -39.1618 |
| R | 0.9533 | 0.9995 | 0.9671 | 0.9997 |

Table.2 Results of robustness measures; BCR

| Attacks | Watermark Only | | Watermark with Encryption | |
|---|---|---|---|---|
| | DCT-DWT | Modified DCT-DWT | DCT-DWT | Modified DCT-DWT |
| Histogram Equalization | 49.5493 | 51.8126 | 71.2354 | 96.2840 |
| Gamma Correction [0.5] | 49.9968 | 63.9981 | 73.5863 | 98.4241 |
| Intensity Adjustment | 25.5869 | 60.0973 | 51.1154 | 98.3722 |
| Blurring [16*16] | 22.9604 | 48.0739 | 52.9086 | 49.7147 |
| JPEG Compression [2:1] | 50.2367 | 50.4086 | 80.1621 | 98.5765 |
| Cropping [50:40] | 25.5869 | 60.0973 | 53.1154 | 98.3722 |
| Mean Filter [16*16] | 22.9767 | 48.1388 | 52.6005 | 49.6920 |
| Median Filter [16*16] | 23.1582 | 48.2685 | 53.1809 | 50.4248 |

## References

[1] S. Lian, "Multimedia content encryption: techniques and applications," CRC Press, 2009, ISBN: 0-8493-8214-9.

[2] B. Schneier, "Applied cryptography", John Wiley & Sons, Inc, 1996, ISBN 0-471-12845-7.

[3] M.Y. Rhee, "Internet security cryptographic principles, algorithms and protocols," Wiley Press, 2003, ISBN-13: 978-0470852859.

[4] J. Seitz, "Digital watermarking for digital media," Information Science Publishing, 2005, ISBN-10: 159140519X.

[5] Y.O. Shi, "Transactions on Data Hiding and Multimedia Security III," Springer, ISBN: 3540690166, 2008.

[6] B. Furht and D. Kiroviski, "Multimedia Security Handbook Internet & Communications," CRC press, 2004, ISBN-10: 0849327733.

[7] A. Haj, "Combined DWT-DCT digital image watermarking," Journal of Computer Science, Vol. 3, No. 9, pp: 740-746, 2007.

[8] F.A. Petitcolas, "Watermarking Schemes Evaluation," IEEE Signal Processing Magazine, Vol.17, pp: 58-64, September 2000.

[9] J.R. Hernández, M. Amado, and F. Pérez-González," DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," IEEE Trans. on Image Processing, Vol. 9, No.1, pp: 55-68, January 2000.

[10] S. Vural, H. Tomii, and H. Yamauchi, "DWT based robust watermarking embed using CRC-32 techniques," World Academy of Science, Engineering and Technology, Vol. 5, pp: 106-109, 2005.

[11] A. Reddy and B. Chatterji, "A New Wavelet Based Logo-watermarking Scheme," Pattern Recognition Letters, Vol. 26, No. 7, pp: 1019-1027, 2005.

[12] W. Chu, "DCT-based image watermarking using subsampling," IEEE Trans. Multimedia, Vol. 5, No.1, pp: 34-38, 2003.

[13] M. Sharkas, D. ElShafie, and N. Hamdy, "A Dual Digital-Image Watermarking Technique," World Academy of Science, Engineering and Technology, Vol. 5, pp: 136-139, 2005.

[14] S.A. Kasmani and A. Naghsh-Nilchi, "A New Robust Digital Image Watermarking Technique Based on Joint DWT-DCT Transformation," Convergence and Hybrid Information Technology. Third International Conference on ICCIT '08, pp: 539 – 544, 2008.

[15] A.J. Menezes, P.C. Van Oorschot, S.A. Vanstone, "Handbook of applied cryptography, 1996, ISBN: 0-8493-8523-7.

[16] M. Salleh, S. Ibrahim & I.F. Isnin, "Image encryption algorithm based on chaotic mapping," *Jurnal Teknologi*, Vol.39, No. D, pp:1–12, 2003.

[17] G. Jakimoski and L. Kocarev."Chaos and cryptography: block encryption ciphers based on chaotic maps". IEEE Trans. on Circuits And Systems, Vol. 48, No.2, pp:163-169, 2001.

[18] S. Li, et. al. "Chaotic encryption scheme for real-time digital video". Proc.of SPI,E Vol. 4666: pp: 149-160, Real-Time Imaging VI, Nasser Kehtarnavaz; Ed., 2002.

[19] P. Junod and S. Vaudenay, "FOX: A new family of block ciphers," in *Proceedings of the SAC 2004 workshop*, 2004, pp. 114–129.

[20] J. Wolkerstorfer, E. Oswald, and M. Lamberger, "An ASIC implementation of the AES SBoxes," in *Proceedings of CT-RSA*, 2002, pp. 67–78.

**Dr. Mohamed Abdel-Azim** received the PhD degree in Electronics and Communications Engineering from the Faculty of Engineering-Mansoura University-Egypt by 2006. After that he worked as an assistant professor at the electronics & communications engineering department until now.

He has 27 publications in various international journal and conferences. His current research interests are in multimedia processing, wireless communication systems, and field programmable gate array (FPGA) applications.

I am a teaching Assistant at the Department of Electronics and communications Engineering, Faculty of Engineering, Mansoura University, Egypt. I got the B. Sc in electronics and communications Engineering 2008; excellent with honor, Faculty of Engineering-Mansoura University.

My areas of interest are security systems and multimedia encryption techniques for cellular communications systems and internet content. I hope to join a research team in one of foreign universities in order to complete my research in this area.


Fig.9 The watermark data

Fig.10 The host image

Fig.11 Watermarked data without encryption


Fig.12 Recovered watermark from Fig.11 encryption

Fig.13 Encrypted original watermark

Fig.14 Watermarked data with encryption

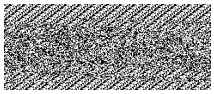


Fig.15 Recovered encrypted watermark


Fig.16 Decryption of original watermark

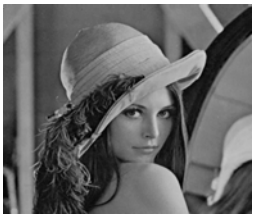


Fig.17 Modified watermark without encryption


Fig.18 Recovered watermark data

Fig.19 Encrypted original watermark

Fig.20 Modified watermark with encryption

Fig.21 Recovered encrypted watermark data


Fig.22 Decrypted data of Fig.21

# A Detailed look of Audio Steganography Techniques using LSB and Genetic Algorithm Approach

**Gunjan Nehru[1], Puja Dhar[2]**

**[1] Department of Information Technology, IEC-Group of Institutions**
**Greater Noida, Uttar Pradesh 201308, India**

**[2]Department of Information Technology, I.T.S-Management & IT Institute**
**Ghaziabad, Uttar Pradesh 201007, India**

**Abstract**

This paper is the study of various techniques of audio steganography using different algorithmis like genetic algorithm approach and LSB approach. We have tried some approaches that helps in audio steganography. As we know it is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. In steganography, the message used to hide secret message is called host message or cover message. Once the contents of the host message or cover message are modified, the resultant message is known as stego message. In other words, stego message is combination of host message and secret message. Audio steganography requires a text or audio secret message to be embedded within a cover audio message. Due to availability of redundancy, the cover audio message before steganography, stego message after steganography remains same. for information hiding.

***Keywords:*** *Audio data hiding, phase coding, LSB,HAS*

## 1. Introduction

The fast improvement of the Internet and the digital information revolution caused major changes in the overall culture. Flexible and simple-to-use software and decreasing prices of digital devices (e.g. portable CD and mp3players, DVD players, CD and DVD recorders, laptops, PDAs) have made it feasible for consumers from all over the world to create, edit and exchange multimedia data. Broadband Internet connections almost an errorless transmission of data helps people to distribute large multimedia files and make identical digital copies of them. In modern communication system Data Hiding is most essential for Network Security issue. Sending sensitive messages and files over the Internet are transmitted in an unsecured form but everyone has got something to keep in secret. Audio data hiding method is one of the most effective ways to protect your privacy.

Encoding secret messages in audio is the most challenging technique to use when dealing with Steganography. This is because the human auditory system (HAS) has such a dynamic range that it can listen over. To put this in perspective, the (HAS) perceives over a range of power greater than one million to one and a range of frequencies greater than one thousand to one making it extremely hard to add or remove data from the original data structure.The only weakness in the (HAS) comes at trying to differentiate sounds (loud sounds drown out quiet sounds) and this is what must be exploited to encode secret messages in audio without being detected.

There are two concepts to consider before choosing an encoding technique for audio. They are the digital format of the audio and the transmission medium of the audio. There are three main digital audio formats typically in use. They are Sample Quantization, Temporal Sampling Rate and Perceptual Sampling.

Sample Quantization which is a 16-bit linear sampling architecture used by popular audio formats such as (.WAV and. AIFF).

Temporal Sampling Rate uses selectable frequencies (in the KHz) to sample the audio. The last audio format is Perceptual Sampling. This format changes the statistics of the audio drastically by encoding only the parts the listener perceives thus maintaining the sound but changing the signal. This format is used by the most popular digital audio on the Internet today in ISO MPEG (MP3). Transmission medium (path the audio takes from sender to receiver) must also be considered when encoding secret messages in audio.

## 2. Methods of hiding various types of information

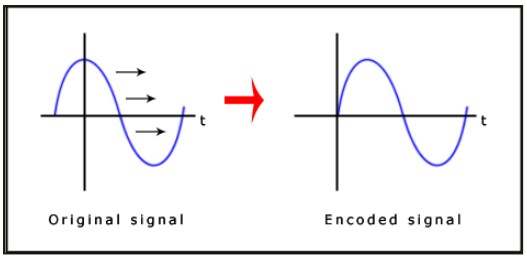
There are numerous methods used to hide information inside of Picture, Audio and Video files. The two most common methods are phase coding, LSB.

### 2.1 Phase coding

Phase coding substitutes the phase of an initial audio segment with a reference phase that represents the hidden data. This can be thought of, as sort of an encryption for the audio signal by using what is known as Discrete Fourier Transform (DFT), which is nothing more than a transformation algorithm for the audio signal.

Human Auditory System (HAS) can't recognize the phase change in audio signal as easy it can recognize noise in the signal. The phase coding method exploits this fact. This technique encodes the secret message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to- noise ratio.

Phase coding addresses the disadvantages of the noise-inducing methods of audio steganography. Phase coding relies on the fact that the phase components of sound are not as perceptible to the human ear as noise is. Rather than introducing perturbations, the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio.




Phase coding is explained in the following procedure:

- The original sound signal is broken up into smaller segments whose lengths equal the size of the message to be encoded.
- A Discrete Fourier Transform (DFT) is applied to each segment to create a matrix of the phases and Fourier transform magnitudes.
- Phase differences between adjacent segments are calculated.
- Phase shifts between consecutive segments are easily detected. In other words, the absolute phases of the segments can be changed but the relative phase differences between adjacent segments must be preserved. Therefore the secret message is only inserted in the phase vector of the first signal segment as follows:

$$phase\_new = \begin{cases} \pi/2 & if \ \ message \ \ bit = 0 \\ -\pi/2 & if \ \ message \ \ bit = 1 \end{cases}$$

- A new phase matrix is created using the new phase of the first segment and the original phase differences.
- Using the new phase matrix and original magnitude matrix, the sound signal is reconstructed by applying the inverse DFT and then concatenating the sound segments back together.

To extract the secret message from the sound file, the receiver must know the segment length. The receiver can then use the DFT to get the phases and extract the information.

One disadvantage associated with phase coding is a low data transmission rate due to the fact that the secret message is encoded in the first signal segment only. This might be addressed by increasing the length of the signal segment. However, this would change phase relations between each frequency component of the segment more drastically, making the encoding easier to detect. As a result, the phase coding method is used when only a small amount of data, such as a watermark, needs to be concealed.

### 2.2 Least Significant bit

Least significant bit (LSB) coding is the simplest way to embed information in a digital audio file. By substituting the least significant bit of each sampling point with a binary message, LSB coding allows for a large amount of data to be encoded. Among many different data hiding techniques proposed to embed secret message within audio file, the LSB data hiding technique is one of the simplest methods for inserting data into digital signals in noise free environments, which merely embeds secret message-bits in a subset of the LSB planes of the audio stream.

This proposed system is to provide a good, efficient method for hiding the data from hackers and sent to the destination in a safe manner. This proposed system will not change the size of the file even after encoding and also suitable for any type of audio file format.
Encryption and Decryption techniques have been used to make the security system robust. Low-bit encoding embeds secret data into the least significant bit (LSB) of the audio file. The channel capacity is 1KB per second per kilohertz (44 kbps for a 44 KHz sampled sequence). This method is easy to incorporate.

## 3. LSB Coding

Sampling technique followed by Quantization converts analog audio signal to digital binary sequence

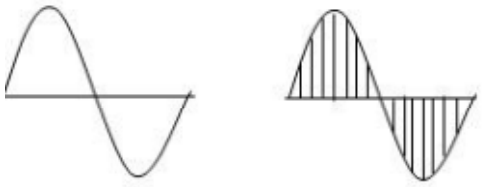
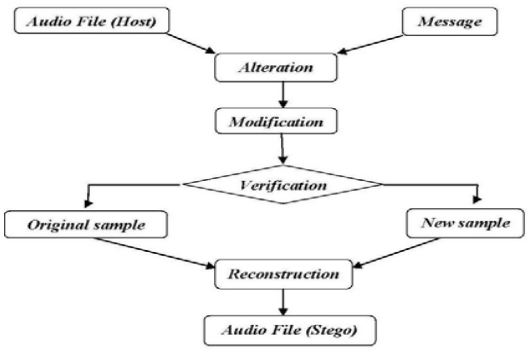


Fig. 3 Sampling of the Sine Wave followed by Quantization process.

In this technique LSB of binary sequence of each sample of digitized audio file is replaced with binary equivalent of secret message. For example if we want to hide the letter 'A' (binary equivalent 01100101) to an digitized audio file where each sample is represented with 16 bits, then LSB of 8 consecutive samples (each of 16 bit size) is replaced with each bit of binary equivalent of the letter 'A'.

## 4. Genetic Algorithm Approach




According to the figure above it shows, there are four main steps in this algorithm.

### 4.1 Alteration

At the first step, message bits substitute with the target bits of samples. Target bits are those bits which place at the layer that we want to alter. This is done by a simple substitution that does not need adjustability of result be measured.

### 4.2 Modification

In fact this step is the most important and essential part of algorithm. All results and achievements that we expect are depending on this step. Efficient and intelligent algorithms are useful here. In this stage algorithm tries to decrease the amount of error and improve the transparency. For doing this World Academy of Science, Engineering and Technolog stage, two different algorithms will be used.
One of them that is more simple likes to ordinary techniques, but in aspect of perspicacity will be more efficient to modify the bits of samples better. Since transparency is simply the difference between original sample and modified sample, with a more intelligent algorithm, I will try to modify and adjust more bits and samples than some previous algorithms. If we can decrease the difference of them, transparency will be improved. There are two example of adjusting for expected intelligent algorithm below.
Sample bits are: 00101111 = 47
Target layer is 5, and message bit is 1
Without adjusting: 00111111 = 63 (difference is 16)
After adjusting: 00110000 = 48 (difference will be 1 for 1 bit embedding) Sample bits are: 00100111 = 39
Target layers are 4&5, and message bits are 11
Without adjusting: 00111111 = 63 (difference is 24)
After adjusting: 00011111 = 31 (difference will be 8 for 2 bits embedding)

## 5. Verification

In fact this stage is quality controller. What the algorithm could do has been done, and now the outcome must be verified. If the difference between original sample and new sample is acceptable and reasonable, the new sample will be accepted; otherwise it will be rejected and original sample will be used in reconstructing the new audio file instead of that Fig. 1 Approach Diagram

## 6. Reconstruction

The last step is new audio file creation. This is done sample by sample. There are two states at the input of this step. Either modified sample is input or the original sample that is the same with host audio file. It is why we can claim the algorithm does not alter all samples or predictable samples. That means whether which sample will be used and modified is depending on the status of samples (Environment) and the decision of intelligent algorithm.

## 7. Literature Review and future scope

In today's world, we often listen a popular term "Hacking". Hacking is nothing but an unauthorized access of data which can be collected at the time of data transmission. With respect to steganography this problem is often taken as Steganalysis . Steganalysis is a process in which a steganalyzer cracks the cover object to get the hidden data. So, whatever be he technique will be developed in future, degree of security related with that has to be kept in mind.
The majority of today's steganographic systems uses multimedia objects like image, audio, video etc as cover media because people often transmit digital pictures over email and other Internet communication.

In a computer-based audio steganography system, secret messages are embedded in digital sound. The secret message is embedded by slightly altering the binary sequence of a sound file. Existing audio steganography software can embed messages in WAV, AU, and even MP3 sound files [22]. Embedding secret messages in digital sound is usually a more difficult process than embedding messages in other media, such as digital images. In order to conceal secret messages successfully, a variety of methods for embedding information in digital audio have been introduced. These methods range from rather simple algorithms that insert information in the similar to watermarks on actual paper and are sometimes used as digital watermarks. Masking images entails changing the luminance of the masked area. The smaller the luminance change, the less of a chance that it can be detected.

Masking is more robust than LSB insertion with respect to compression, cropping, and some image processing. Masking techniques embed information in significant areas so that the hidden message is more integral to the cover image than just hiding it in the "noise" level. This makes it more suitable than LSB with, for instance, lossy JPEG images. We use a more powerful GA (Genetic Algorithm) based LSB (Least Significant Bit) Algorithm to encode the encrypted message into audio data. Here encrypted message bits are embedded into random and higher LSB layers, resulting in increased robustness against noise addition. On the other hand, GA operators are used to reduce the distortion. Using the proposed genetic algorithm, message bits are embedded into multiple, vague and higher LSB layers, resulting in increased robustness. The robustness specially would be increased against those intentional attacks which try to reveal the hidden message and also some unintentional attacks like noise addition as well.

## 8. Conclusion

In this paper we have introduced a robust method of imperceptible audio data hiding. This system is to provide a good, efficient method for hiding the data from hackers and sent to the destination in a safe manner. This proposed system will not change the size of the file even after encoding and also suitable for any type of audio file format. Thus we conclude that audio data hiding techniques can be used for a number of purposes other than covert communication or deniable data storage , information tracing and finger printing, tamper detection. As the sky is not limit so is not for the development. Man is now pushing away its own boundaries to make every thought possible. So similarly these operations described above can be further modified as it is in the world of Information Technology.

## References

[`1] Cvejic N. and Seppänen T. "Increasing the capacity of LSB based audio steganography", Proc. 5th IEEE International Workshop on Multimedia Signal Processing, St. Thomas, VI, December 2002, pp. 336-338.

[2] Westfeld A. and Pitzmann A. "Attacks on Steganographic Systems". Lecture Notes in Computer Science, vol. 1768, Springer-Verlag, Berlin, pp. 61-75, 2000.

[3] Fridrich, Jessica and others. "Steganalysis of LSB Encoding in Color Images." Proceedings of the IEEE International Conference on Multimedia

**First Author** She is having experience of nine years in academics. She is working as Assistant Professor in Department of Information Technology at I.E.C, She has done M.Tech(IT).

**Second Author** She is having experience of more than six years in academics. She is working as Assistant Professor in Department of Information Technology at I.T.S Mohan Nagar, She has published number of papers in proceedings

of national and international conferences and refereed journals. She has done M.Tech(IT),Msc(IT). She has received 'Best Faculty Award' at I.T.S. Her research interest includes networking and databases.

# Design of Power Efficient Schema for Energy Optimization in Data Center With Massive Task Execution Using DVFS

Arunadevi.M[a], Dr. R.S.D Wahidabanub[b]
[a]Assistant Professor of Sambhram Academy of Management Studies, Bangalore, India,
Mobile: 9538202730
[b]Professor and Head of Department of ECE of Government College of Engineering, Salem, India
Mobile: 9443008886

**Abstract: The proposed system highlights a novel energy efficient technique by considering an entire datacenter using DVFS (Dynamic Voltage and Frequency Scaling). Unwanted power consumption was always a matter of concern from last 2 years for the administrators of datacenters. Therefore, a research trial in order to minimize the power consumption has become of the prominent concern in the area of cloud computing as it claims zero down-time as per Service Level Agreement. In this research work, we consider the case of multimedia video sharing web services, which we believe is frequently in top of the hit counter with millions of user sharing high size of video application, giving rise of power consumption from data center. The proposed system presents a unique provisioning technique which amalgamates the power factor with network resources. The simulation results also show that the system is successful in maintaining an equilibrium between power utilization at data center along with unit of task allocation processing.**

**Keywords**: Energy consumption, Cloud Computing, Data center.

## I. INTRODUCTION

The current research concern is the unwanted power utilized in data center which is exceptionally gaining attention of researchers with respect to scheduling of the computing resources. Not only this, the investment towards the services of cloud computing is very high [1]. Cloud computing assures virtually limitless computational resources to its users, while letting them pay only for the resources they actually use at any given time. Various cloud computing services such as Amazon EC2 [2] and Google App Engine [3] are designed to take benefit of the already existing infrastructure of their respective business which delivers computing services to users as a utility in a pay-as-you-go manner [4]. The different services facilitated by the cloud providers are Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and finally Software as a Service (SaaS). In reality, Service providers make high-quality use of IaaS and PaaS for developing their services without consideration of physical hardware, while users also can access on-demand and pay-per-use services anywhere in Cloud computing. But one of major issue in datacenters found is to manage optimum energy usage in the systems. It was found that there is a consumption of 10 to 100 times high energy per square foot compared to distinctive office building for an average scale datacenters [5]. They can even consume as much electricity as a whole city [6]. The majority of energy consumption in datacenters approaches from computation processing, disk storage, network, and various types of cooling systems. In this research proposal, we will deploy virtual machine provisioning to analyze various energy processing functionalities along with their parameters. This can be one of the noteworthy technique in cloud computing for evaluation of energy models in Cloud computing. The concept of cloud computing guarantees Service Level Agreements (SLAs) between customers and Cloud suppliers with Pay-as-you-go mechanism which specify that the negotiated agreements as deadline. Therefore, every datacenters will attempt to minimize energy consumption without violating these SLAs. As various real-time applications will require deadline constraints, this research proposal will concentrate on power-aware real-time Cloud services, such as distributed image processing, realtime distributed databases, financial analysis and so on. The main contribution of this proposed system will be to give (i) analysis and in-depth study of real-time Cloud service with virtualization, and (ii) to probe various energy-aware virtual machine optimizing schemes based on Dynamic Voltage Frequency Scaling schemes (DVFS).

The main contribution of this paper is to highlight a novel provisioning methodology which amalgamates power efficiency along with network resources information. My previous work was much focus on

analyzing power consumption in virtual machine as well as processing element with respect to SLA (Service Level Agreement). The work has also shown the comparative analysis of implication of DVFS as well as DVS (Dynamic Voltage Scaling) technique. This work is more stressed on larger scale i.e. datacenter using similar DVFS. The proposed system targets to maintain the equilibrium between the unit task accomplishment, task quality of server needs, requirements of the network resources, and the utilized power by the data center. It is also found that data with known task will require a less effective load towards the design of queue structure. For the purpose of experiment, a multimedia sharing web application is considered.

The remainder of this paper is organized as follows. Section II presents problem statement followed by related works in Section III. The proposed approach is discussed in Section IV. Section V will highlight the simulation results followed finally Section VI will conclude the research proposal.

## II. PROBLEM DESCRIPTION

Majority of the prior research work done in the area of analyzing power utilization mainly concentrates on task scheduling in the center with respect to task allocation among the application servers [7], targeted power saving [8] or the criteria considering thermal factors [9]. The major research gap is that there are only few implementation work being carried out in past considering data centers, unwanted power consumption along with task provisioning. [10][11].

Cloud computing has been accounted for diversified emerging issues which is yet to be resolved. Unfortunately, the current potentials and capability of cloud computing has yet not been enhanced to cater the crucial scope of the usage. There are technical and nontechnical gaps. One of the most prominent issues is that although vision of cloud computing is to deploy at any location irrespective of time and devices, particularly in case of distributed system and existing clouds is reported to deliver very poor performance in applications. Any significant issue is the migration of data from client's infrastructure to clouds actually cost very high and it is also a time consuming process. Various legal restrictions are yet to be created for the acceptance of this new technology in socioeconomic market. In many data centers today, the cost of power is only second to the actual cost of the equipment investments. Today the cloud data center consumes 1-2 percent of world energy. If left unchecked, they will rapidly consume an ever-increasing amount of power consumption.

## III. RELATED WORK

Suzanne [12] propose and motivate JouleSort, an external sort benchmark, for evaluating the energy efficiency of a wide range of computer systems from clusters to handhelds. Since JouleSort focuses on data management tasks, it misses some important energy-relevant components for multimedia applications. JouleSort omits displays, which are an important component of total power for mobile devices. Anand Vanchi [13] developed methods for identifying measurable efficiency improvements and placed instrumentation to continuously track power usage effectiveness (PUE), the key metric of data center energy efficiency. Anshul [14] presents a novel approach to correctly allocate resources in data centers, such that SLA violations and energy consumption are minimized. Dara Kusic et. al. [15] implement and validate a dynamic resource provisioning framework for virtualized server environments wherein the provisioning problem is posed as one of sequential optimization under uncertainty and solved using a look ahead control scheme. A new suite for placement and energy consolidation of the virtual machines in data centers has been discussed by Michael Cardosa [16]. The author has validated the results obtained from experiments conducted in artificial and real data centers testbeds and concluded that the protocol designed constantly provides the optimal throughput on a large spectrum of inputs. A unique design and accomplishment of an architecture for managing resources over the hosting center operating system has been proposed by Jeffrey S. Chase [17]. The prime importance of the work was the power for running the resource management problems for huge clusters of servers intended to provide automation in adapting to prescribed load for server resources. This has resulted in enhancement of the power efficiency for clusters of server by resizing the set of active servers dynamically and reacting to the supply of energy interference or thermal events by corrupting the service as per Service Level Agreements (SLAs).

A power-aware protocol which automatically acclimatizes its potential difference and frequency configuration is already proposed by Chunghsing Hsu [18]. The main intention is to facilitate considerable reduction in energy and power cutback with optimal effect on its respective performance. Kyong Hoon Kim [19] proposed power-aware

scheduling algorithms for bag-of-tasks applications with deadline constraints on DVS-enabled cluster systems. The proposed scheduling algorithms select appropriate supply voltages of processing elements to minimize energy consumption. Linwei Niu [20] investigated the problem of applying scheduling techniques to reduce both the dynamic and leakage energy consumption. Monfort [21] proposed a hierarchical real-time virtual resource model which is ideal for the open system environment with a clean separation of concerns between task group scheduling and partition scheduling across multiple levels of resource decomposition. Anshul Gandhi [22] experimentally find that the power-to-frequency relationship within a server for a given workload can be either linear or cubic. Interestingly, we see that the relationship is linear for DFS and DVFS when the workload is CPU bound, but cubic when it is more memory bound. By contrast, the relationship for DVFS+DFS is always cubic in their experiments. Leping Wang [23] presents a threshold-based method for efficient power management of heterogeneous soft real-time clusters. An illustrative analysis for recognizing the environment for isolative applications was studied by Akshat Verma [24]. The author has also shown that energy utilization by various HPC applications can be dependent on application or non-linear and can posses huge diversified range. The inter-association of power utilization and workload performance has been analyzed by Shekhar Srikantaiah [25] which elicited the power performance trade-offs for consolidation and concludes that superior operating point do persists.

## IV. PROPOSED SYSTEM

The proposed technique attempts to reduce the cumulative power utilization of a specified data center which is done by choosing the suitable evaluating resources for task processing depending on the load quantity and transmission capability of the variants of data center specified. The transmission capability will be represented by available bandwidth facilitated to the unique application servers or clusters of servers. The proposed technique designs a architectural framework possessing the better designed topology of data center. The weighted grouping of the server level $W_s$, rack level $W_r$, and module level $W_m$ is given by,

Weighted Grouping = P .Ws + Q. $W_r$ + R. $W_m$. (1)

Where P, Q, and R are the weighted coefficients factors. The allocation of higher set of load towards the server is preferred if the value of P is very high. Similarly, priority will be designed higher for higher set of task in the racks with reduced traffic events when the value of Q is higher. The selection of loaded modules will be preferred if value of R is higher. For understanding the magnitude of the task allocation in datacenter, the value of R plays an important role. The implementation plan of the proposed system is as shown in Fig 1.

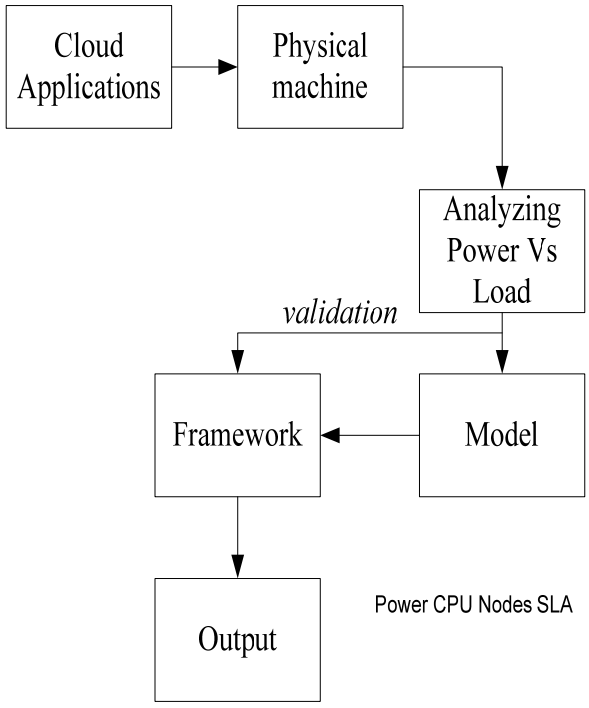



Fig 1. Implementation Plan

The selection of the operationally processing server will group the server load $S_L(l)$ along with its transmission capability $T_C(q)$ which is analogous to equivalent distribute of the uplink resources on the top of rack switch.

$$W_s(l,q) = S_L(l).\frac{T_C(q)^{\varphi}}{\Delta_r} \quad (2)$$

In the above equation, $S_L(l)$ is a parameter which is proportional to the load of any individual servers l, $T_C(q)$ which represents the load at the rack uplink by evaluating the traffic blocking intensity in the designed queue q, $\Delta_r$ is a bandwidth over scheduled parameter at the rack switch and φ will represent a coefficient for share between $S_L(l)$ and $T_C(q)$. It is also known that both $S_L(l)$ and $T_C(q)$ should lie

within a range (0,1) greater value φ will reduce the significance of network variants $T_C(q)$. Therefore, with context to the equation (2), the criteria having influencing the rack switch can be expressed as:

$$W_r(l,q) = S_R(l).\frac{T_m(q)^{\varphi}}{\Delta_m} = \frac{T_m(q)^{\varphi}}{\Delta_m}.\frac{1}{n}\sum_{i=1}^{n} S_L(l) \quad (3)$$

$$W_m(l) = S_m(l) = \frac{1}{k}\sum_{j=0}^{k} S_R(l) \quad (4)$$

In the equation (4), $S_R(l)$ is a load on rack which is derived from the addition of all the component loads on server in the rack, $S_M(l)$, is the module load derived as the normalized addition of all of the loads in rack in this module, n and k are quantity of the servers in a rack and the quantity of the racks in the module respectively, $T_C(q)$ is dependent to the network load at incoming networking devices and $\Delta_m$ will represents channel capacity as over scheduled criteria at the module switches. It can be seen that module level criteria Fm will posses only components *l* related to load. This phenomenon is due to the reason that all the components are actually associated to the uniform networking device and split the similar channel capacity by deploying a technique which can work on unit load balancing resulting in discretion of traffic flows by estimating a hash function on all ingress message. It was also noted that idle server in data center consumes 2/3rd of its crest utilization [26] which concludes that if any provisioner is to be designed that it must merge all the allocated task on a minimum feasible group of computing resources. Also, it should be noted that if the application servers are kept in uniform execution mode at high frequency than hardware reliability along with its performance will be not so much effective. Therefore, the proposed load criteria is designed as following:

$$S_L(l) = \frac{1}{1+e^{-10(l-\frac{1}{2})}} - \frac{1}{1+e^{-\frac{10}{\varepsilon}(l-(1-\frac{\varepsilon}{2}))}} \quad (5)$$

The preliminary component as shown in equation (5) explains the data-structure of the function while the 2nd variant denotes the fining method targeted at the junction towards the highest load of the server.

The factor ε represents the dimension and the shift of the descent slope. The load of application server *l* should be in constraint of (0,1). In case of highly dynamic load towards server, the load on application server will be evaluated as the total of all processing load of the allocated present task. But in case of jobs where the deadline is fixed for completion, the load on the server can be represented as the least quantity of the networking resources which will be required from the application server to accomplish all the jobs at pre-defined time as per SLA. In datacenter, the application server will split the top of rack switch for meeting their transmission requirements. Conversely, representing a section of this channel capacity utilized by a specified application server or a transmission at higher frequency deployed in defined server during execution will be definitely highly expensive in terms of processing the algorithm. In order to mitigate such issues, using eq 3 and 4 will use a variant based on the task allocated stage of the job-queue at the networking device and levels the channel capacity over scheduling criteria Δ.

As an alternative of depending on the appropriate queue size, the allocation stage q will be leveled with the cumulative queue size $T_{max}$ within the range of (0,1). The variety is analogous to void and complete buffer allocation. By depending on the buffer allocation, the proposed system will be responsive to the increasing blocking in the racks moderately in comparison to rate of transmission differences. Therefore T(q) will be represented as:

$$T(q) = e^{-(\frac{2q}{T_{max}})^2} \quad (6)$$

## V. SIMULATION RESULT

The proposed system discussed is evaluated on java platform in 32 bit machine with windows OS of 1.84 GHz processor. The prototype design of DVFS protocol is layered into task-allocation, group creation, and processing elements. The virtual machines are developed for accepting number of task considering simulation time and processing speed of the system. The framework is developed to track down the communication in datacenter. The optimization is checked through JOULEX [27]. The application server is taken for YOUTUBE [28] as normally multimedia applications are heavy, frequently visited, and has sharing privileges for the users..

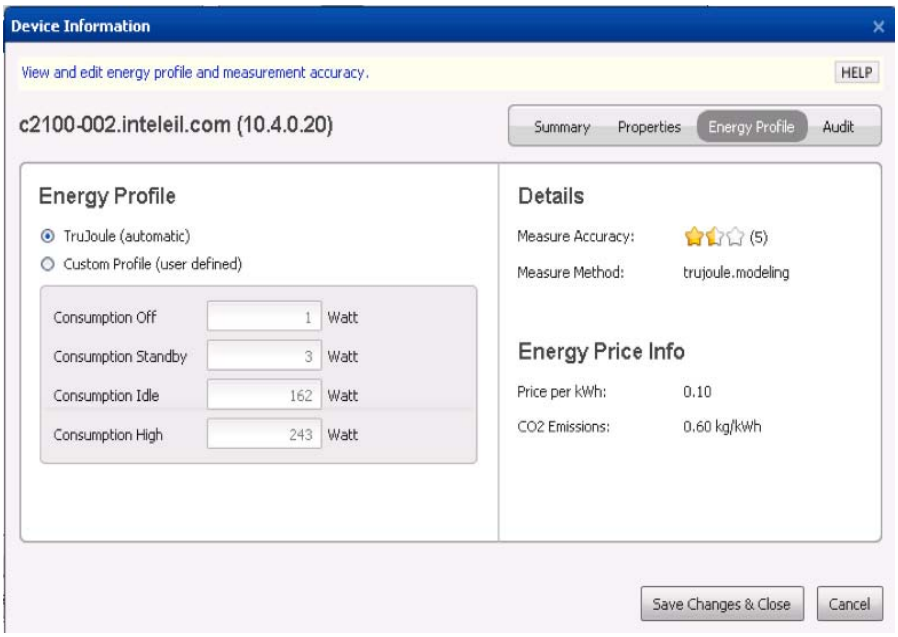



Fig 2. Joulex Interface

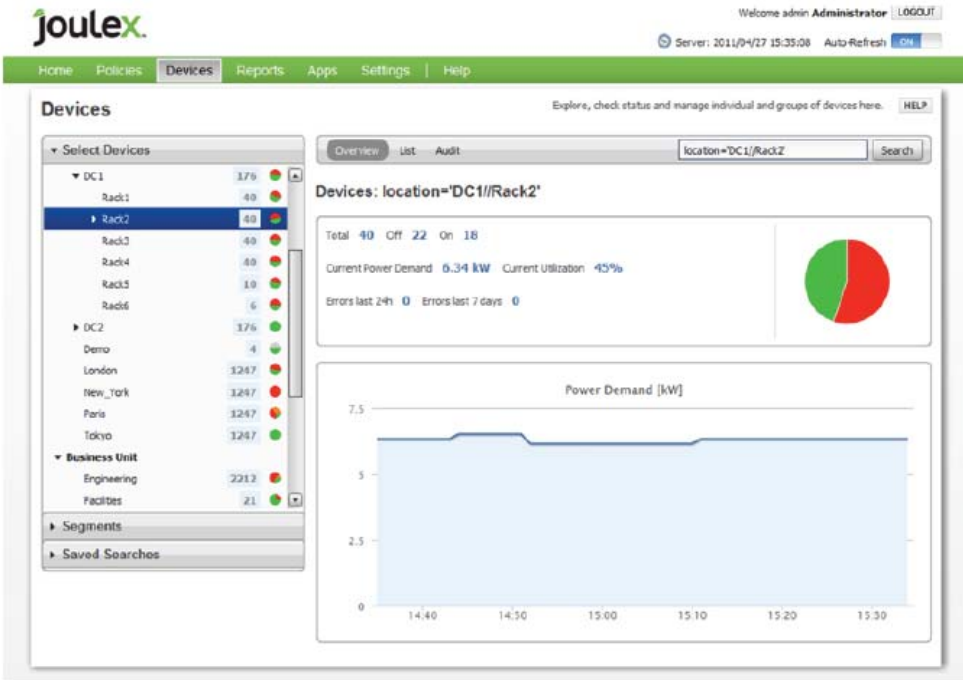



Fig 3. Joulex interface with graphical representation of Power consumption

The proposed system performs implementation of our proposed DVFS scheme along with traditional DVS scheme, where we consider both the schemes for energy optimization for the datacenters along with all its components. The considerations are:

No of server=1536
No. of Racks=32
No of application server by racks=48
Propagation delay on links=10 ns
Experiment Task size=15 KB

A communication link is created for connecting application server within the rack. The total set of work is exponentially dispersed to the executers.

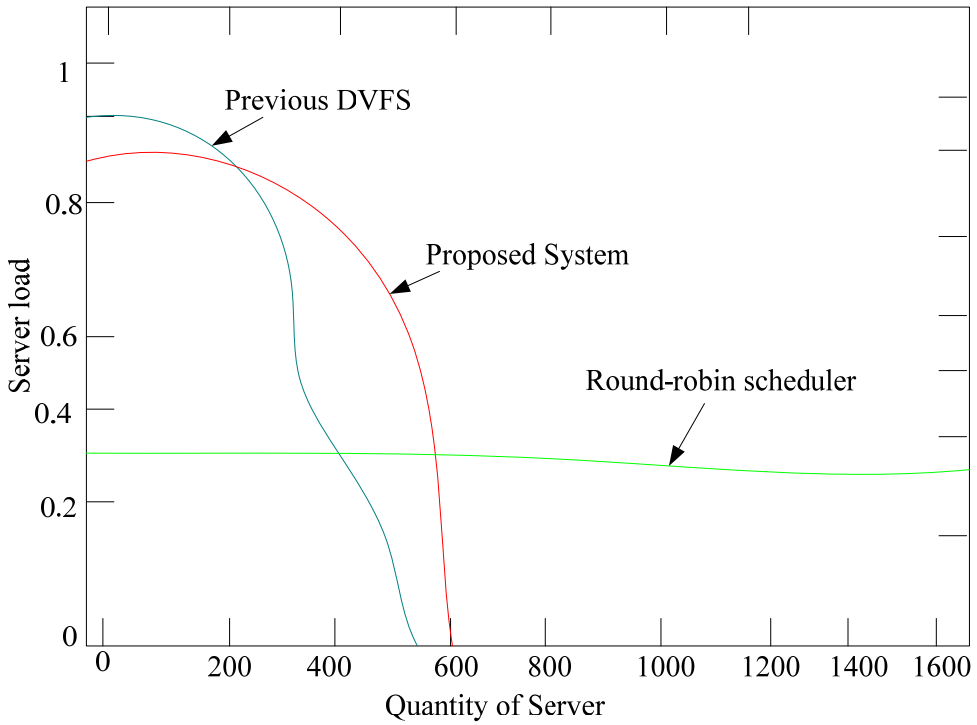



Figure 4. Server workload distribution performed by proposed system, previous DVFS implementation, and round-robin technique

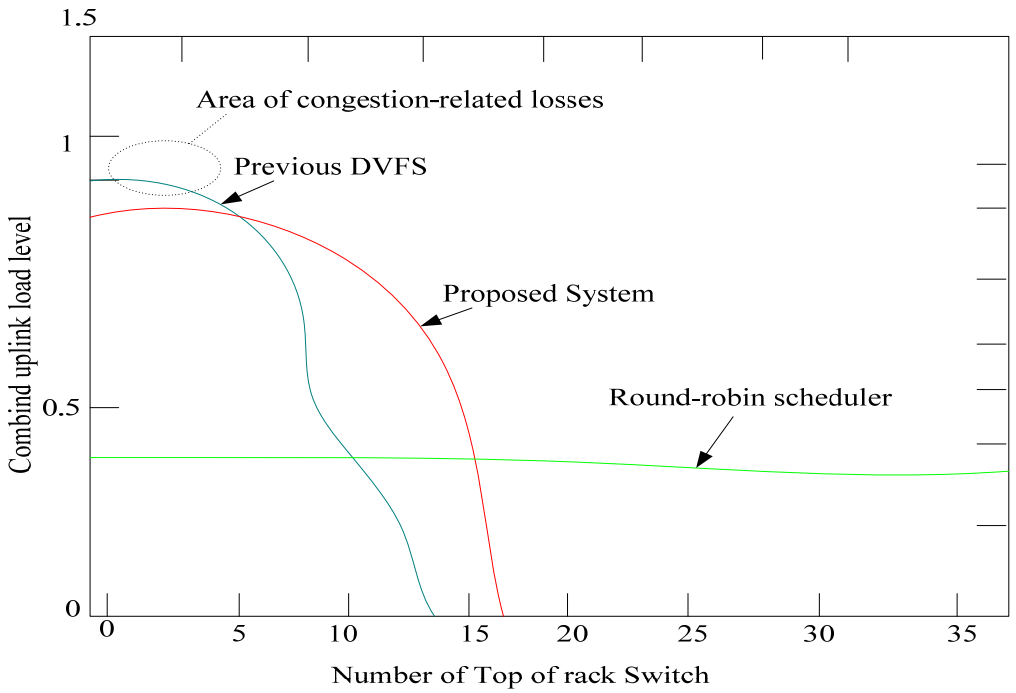



Figure 5. Joint uplink traffic load

Once the provisioning decision is undertaken for a recently designed task, it will be forwarded to the

point of datacenter to the chosen application server for the purpose of execution. At the processing mode, the task generates a uniform stream of bitrate of 1.6 Mb/s which is forwarded to the datacenter.

To increase the challenges at the simulation time, each of the task communicates with another task which is arbitrarily chosen by transmitting a dummy message of 125 KB implicitly. Also at the event of task accomplishment, the similar size packets will be forwarded to the datacenters. The mean task load in the datacenter is fixed to 45% which is dispersed among other application server. A round robin algorithm can be used for this purpose.

The proposed algorithm is experimented with the previously implemented DVFS techniques as well as conventional round robin algorithm. The comparative analysis is presented in fig 4 and fig 5. Fig. 4 represents the load allocation to Server for all the analyzed provisioner and Fig 5 represents a Joint uplink traffic load.

## VI. CONCLUSIONS

The proposed system identifies the function of transmission structure in the datacenter power utilization and highlights a unique process which amalgamates the power with network resources. The proposed system maintains the equilibrium for the unique task processing and accomplishment, power utilization of the data center, and optimal network requirements. The system has optimized the research gap for task grouping for reducing the quantity of the application server in the data center and dispersing of the network resources in order to mitigate the congestion which might possible occur in data center leading to unwanted power consumption.

**Reference**

[1] R. Brown et al., "Report to congress on server and data center energy efficiency: Public law 109-431," Lawrence Berkeley National Laboratory, 2008.
[2] "Amazon Elastic Compute Cloud," http://aws.amazon.com/ec2.
[3] "Google App Engine," http://code.google.com/appengine/.
[4] M. Armbrust, et al. Above the Clouds: A Berkeley view of cloud computing. Tech. Report No. UCB/EECS-2009-28, University of California at Berkeley, USA, February 2009
[5] P. Scheihing. Creating energy efficient data centers. In Data Center Facilities and Engineering Conference. Washington, DC, USA, May 2007.
[6] J. Markoff and S. Lohr. Intel's huge bet turns iffy. New York Times Technology Section, September 2002.
[7] Ying Song, Hui Wang, Yaqiong Li, Binquan Feng, and Yuzhong Sun, "Multi-Tiered On-Demand Resource Scheduling for VM-Based Data Center," IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID), pp. 148 – 155, May 2009.
[8] A. Beloglazov, and R. Buyya, "Energy Efficient Resource Management in Virtualized Cloud Data Centers," IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid), pp. 826 –831, May 2010.
[9] Qin Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-Efficient Thermal-Aware Task Scheduling for Homogeneous High-Performance Computing Data Centers: A Cyber-Physical Approach," IEEE Transactions on Parallel and Distributed Systems, vol.19, no. 11, pp. 1458 – 1472, November 2008.
[10] M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, and A. Vahdat, "Hedera: Dynamic Flow Scheduling for Data Center Networks," in Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation (NSDI '10), San Jose, CA, April 2010.
[11] Xiaoqiao Meng, V. Pappas, and Li Zhang, "Improving the Scalability of Data Center Networks with Traffic-aware Virtual Machine Placement," IEEE INFOCOM, San Diego, California, March 2010.
[12] Suzanne Rivoire, Mehul A. Shah, Parthasarathy Ranganathan, Christos Kozyrakis, JouleSort: A Balanced Energy-Efficiency Benchmark, Proceedings of the ACM SIGMOD Intl. Conference on Management of Data (SIGMOD), Beijing, China, June 2007.
[13] Anand Vanchi, Sujith Kannan, and Ravi Giri, Increasing Data Center Efficiency through Metering and Monitoring Power Usage, White Paper, 2009
[14] Anshul Gandhi, Yuan Chen, Daniel Gmach, Martin Arlitt, Manish Marwah, Minimizing Data Center SLA Violations and Power Consumption via Hybrid Resource Provisioning, IEEE, 2010
[15] D. Kusic, et al. Power and performance management of virtualized computing environments via lookahead control. In Proc. of 5th IEEE Intl. Conf. on Autonomic Computing (ICAC 2008), pages 3–12. Chicago, USA, June 2008.
[16] M. Cardosa, M. R. Korupolu, and A. Singh. Shares and utilities based power consolidation in virtualized server environments. In Proc. of

IFIP/IEEE Intl. Symp. On Integrated Network Management. USA, June 2009.
[17] J. S. Chase, et al. Managing energy and server resources in hosting centers. In Proc. of 8th ACM Symp. on Operating Systems Principles. Banff, Canada, October 2001.
[18] C. Hsu and W. Feng. A power-aware run-time system for high-performance computing. In Proc. of ACM/IEEE Conf. on Supercomputing. Seattle, USA, November 2005.
[19] K. H. Kim, R. Buyya, and J. Kim. Power aware scheduling of bag-of-tasks applications with deadline constraints on DVS-enabled clusters. In Proc. of 7th IEEE Intl. Symp. On Cluster Computing and the Grid (CCGrid'07), pages 541–548. Rio de Janeiro, Brazil, May 2007.
[20] L. Niu and G. Quan. Reducing both dynamic and leakage energy consumption for hard real-time systems. In Proc. Of CASES'04. Washington, DC, USA, Sept. 2004
[21] Valérie Monfort, Maha Khemaja, Nouha Ammari, Fayssal Fehli, Using SaaS and Cloud computing For "On Demand" E Learning Services, 2010 10th IEEE International Conference on Advanced Learning Technologies
[22] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy. Optimal power allocation in server farms. In Proc. of Intl. Joint Conf. on Measurement and Modeling of Computer Systems, pages 157–168. Seattle, USA, June 2009.
[23] L. Wang and Y. Lu. Efficient power management of heterogeneous soft real-time clusters. In Proc. of IEEE Real-Time Systems Sym. Barcelona, Spain, Dec. 2008.
[24] A. Verma, P. Ahuja, and A. Neogi. Power-aware dynamic placement of HPC applications. In Proc. of ICS'08, pages 175–184. Agean Sea, Greece, June 2008.
[25] S. Srikantaiah, A. Kansal, and F. Zhao. Energy aware consolidation for cloud computing. In Workshop on Power Aware Computing and Systems (HotPower '08). San Diego, USA, December 2008.
[26] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-aware server provisioning and load dispatching for connection intensive internet services," the 5th USENIX Symposium on Networked Systems Design and Implementation, Berkeley, CA, USA, 2008.
[27] http://www.joulex.net/.
[28] http://www.youtube.com

# Helpful Business Value of Advance Bal Information System

**Muhammad Awais (*Assistant Professor*)**
**Department of Computer Science, NFC Institute of Engineering & Fertilizer Research, Faisalabad**
**Faisalabad, 38000, Pakistan**

**Muhammad Irfan (*Lecturer*)**
**Department of Statistics, Government College University, Faisalabad**
**Faisalabad, 38000, Pakistan**

**Muhammad Bilal**
**University of Engineering & Technology Lahore**
**Faisalabad, 38000, Pakistan**

**Tanzila Samin *(Lecturer)***
**School of Business Management, NFC Institute of Engineering & Fertilizer Research, Faisalabad**
**Faisalabad, 38000, Pakistan**

***Abstract***

It has become progressively more complicated to pay more attention to the significance of business value of information system in early days. Above the past few days researchers have high bleached the need of information system in business field. Almost every organization like electronics, textile, computer, etc is investing significantly in information system. It is generally observed that IS savings facilitate firms to achieve competitive benefit and enhance central part competencies to improve their performance and increase further earnings.

As a result Helpful Business Value of Advance BAL Information System (HBVABIS) is at the heart of our thoughtfulness in this research paper. This paper seeks to address the following questions. What is meant by business value of information system? How can we calculate business value of information system? What is the importance of information system in new e-commerce era? It has been generally observed that with the passage of time as the quick changes are taking place in information system the above questions are gaining importance in the field of business. The consideration in proper selection of technology and its proper exploitation to improve the performance of a business in education sector help to attain the business value. We also covered some aspects of educational sector using statistical analysis.

***Keywords:*** -Helpful Business Value of Advance BAL Information System (HBVABIS), Performance, e- commerce, statistical analysis etc.

## 1. Introduction

Basically IS (Information System) is any combination of information technology and people's actions using that technology to support operations, administration. It is also called application landscape. Simply information system is commonly used to refer the interaction among people, algorithmic processes, data and technology. [1]

An information system can also be considered a semi-formal language which supports human decision making and action. An information system (IS) is a work system whose actions are dedicated to processing (capturing, transmitting, storing, retrieving, manipulating and displaying) information. [1]

### 1.1 Business Value

What is Business Value? Business value is used as middleware in Business Goals and Information Technology. Basically Business value is of vital importance among these parameters. The given Fig.1 shows actual structure.

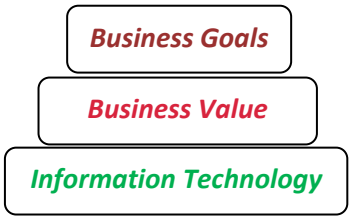



Fig. 1 [3]

### 1.2 Information System Types

Information System is generally classified into five categories:-

- Office Information System (OIS)
- Transaction Processing System (TPS)
- Management Information System (MIS)
- Decision Support System (DSS)
- Executive System (ES)

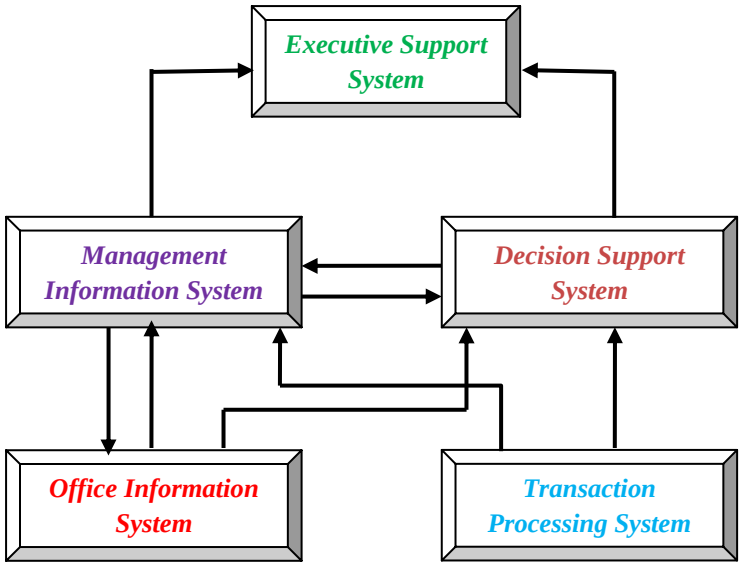



Fig. 2

**Office Information System**

Office Information System (OIS) is an information system that uses hardware, software and networks to enhance work flow and make easy communication among employees. Win an office information system, also described as office automation; employees execute tasks electronically using computers and other electronic devices, instead of manually. With an office information system, for example, a registration department might post the class schedule on the Internet and e-mail students when the schedule is well-run. In a manual system, the registration department would photocopy the schedule and mail it to each student's house.[2]

**Transaction Processing System**

A Transaction Processing System (TPS) is an information system that captures and processes data generated during an organization's everyday transactions. A transaction is a business activity such as a deposit, payment, order or reservation. For example clerical staff typically performs this activity.[4]

**Management Information System**

A Management Information System (MIS) is an information system that generates accurate, timely and organized information so managers and other users can make decisions, solve troubles, administer activities, and track progress. Because it generates reports on regular basis, a management information system sometimes is called a management reporting system (MRS). MIS generates three basic types of information: detailed, summary and exception. Example of a detail report is Detailed Order Report. Example of a summary report is Inventory Summary Report. Example of an exception report is an Inventory Exception Report. [2]

**Decision Support System**

A Decision Support System (DSS) is an information system designed to help users attain a decision when a decision-making situation arises. A variety of DSS's exist to help with a range of decisions. A decision support system uses data from internal and/or external sources. Internal sources of data might include sales, manufacturing, inventory, or financial data from an organization's database. Data from external sources could include interest rates, population trends, and costs of new house construction or raw material pricing. [2]

**Expert System**

An Expert System (ES) is an information system that captures and stores the knowledge of human experts and then imitates human reasoning and decision-making processes for those who have less expertise. Expert Systems are composed of two components.

- knowledge base
- Inference rules

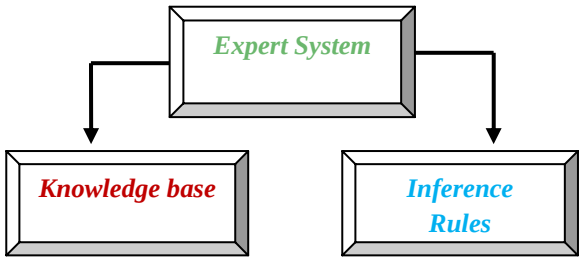



Fig. 3

## 2. Available Solution

In this available solution the **Perceived Value Approach** is compared with Normative Value approach and Real Value Approach.

The Perceived Approach is based on subjective evaluations performed by users of an information system. Similar to the Real Approach, empirical research methods are used to monitor and control implemented or prototyped systems to review their potential impacts (usually via some type of survey tool). This approach is well-suited for examination of information security value issues where risks are uncertain

The Perceived Approach is not without its weaknesses in applicability. According to Ahituv (1989), the Perceived measure has several problems for quantifying information value:

- Point of measurement. The Perceived Approaches examine the outcomes generated by the decision maker rather than the outcome generated by the system. While this does separate Real from Perceived, a user might believe a system to be good, whereas in a Real sense another system might be better (i.e.an insecure system may be easier to log into, therefore perceived superior to a more secure system).
- Voluntary system. One must have subjects to survey who are qualified to make these judgments.

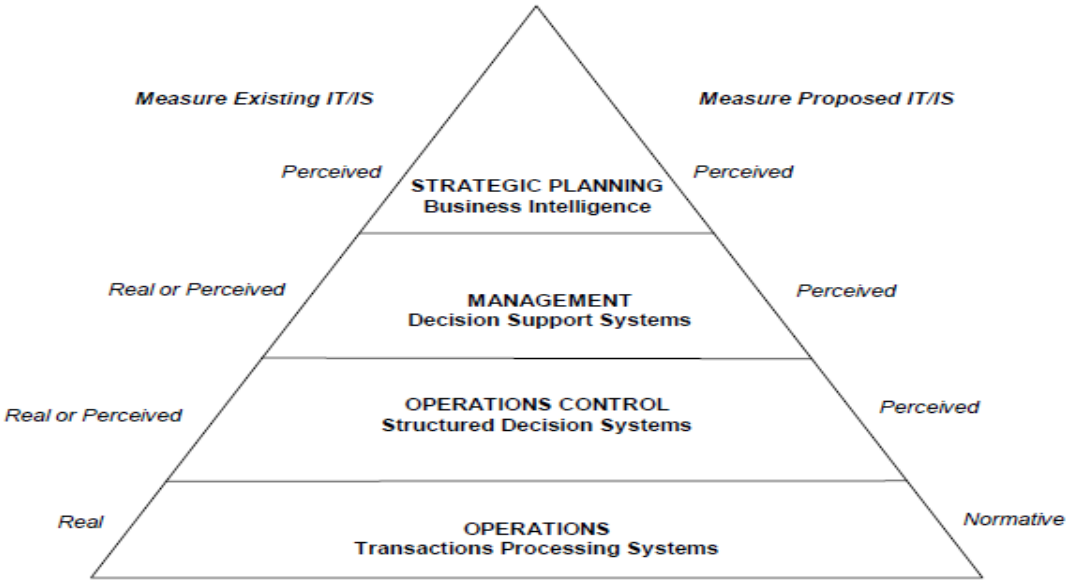



Fig. 4[3]

The Perceived Value Approach is better than others two Approaches as comparing in Fig. 4. But in Perceived Value Approach risk factor is occurred. And Perceived Value Approach is only applicable within the organization.

## 3. Methodologies

There are some problems that exist in available solutions. Helpful Business Value of Advance BAL Information System (HBVABIS) will solve these problems.

In Helpful Business Value of Advance BAL Information System (HBVABIS), we analyzed that University System is better than College System in education sector. We proved it with a solid reason and through proper statistical analysis. Because our aim is to achieve the effective business value (business value lie in between business goals and information technology) in information system where Information Technology is used as business partner. In HBVABIS, We focused on the following information system categories:

- Management Information System (MIS)
- Decision Support System (DSS)
- Transaction Processing System (TPS)
- Expert System (ES)

### 3.1 Comparative Analysis & Discussion

We compared the previous available solutions with Helpful Business Value of Advance BAL Information System (HBVABIS).

In the Previous **Perceived Value Approach** solution implemented only within the organization. In other word, simply this previous solution is applicable to centralized structure, and few information system types are used.

Helpful Business Value of Advance BAL Information System **(HBVABIS)** is comparatively better than the existing available solution. We focused the College System and University system in three information categories. We explained with the structural diagram. We highlighted the advantages of University System as compare to College System in this diagram. We used four mention information system categories like MIS, DSS, TPS and ES in University System. The above mentioned four system categories are lacking in College System.

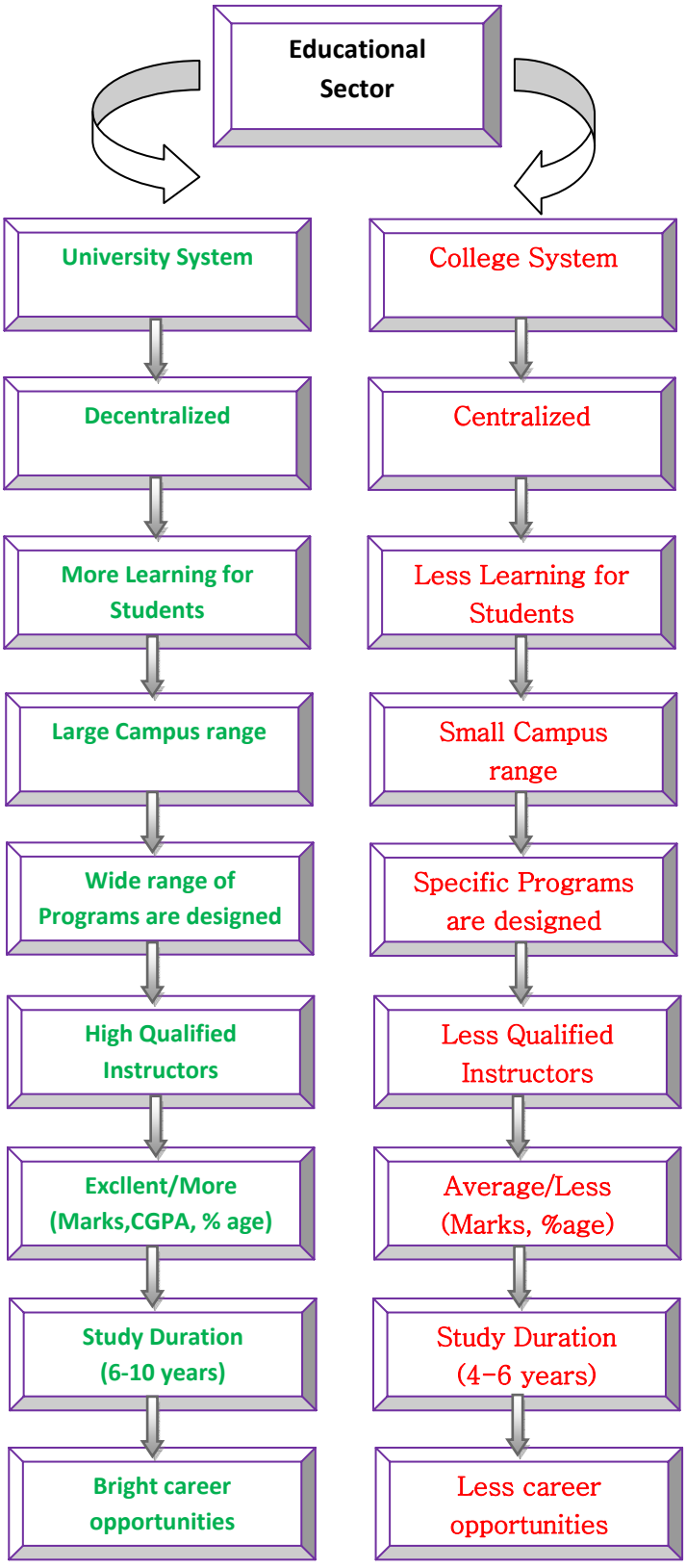



Fig. 5

In the given Fig. 5, we explained that University system is better than College system. Because, University System consists of long time duration for students study (Maximum Ph.D level) but College System has short time duration for students study.

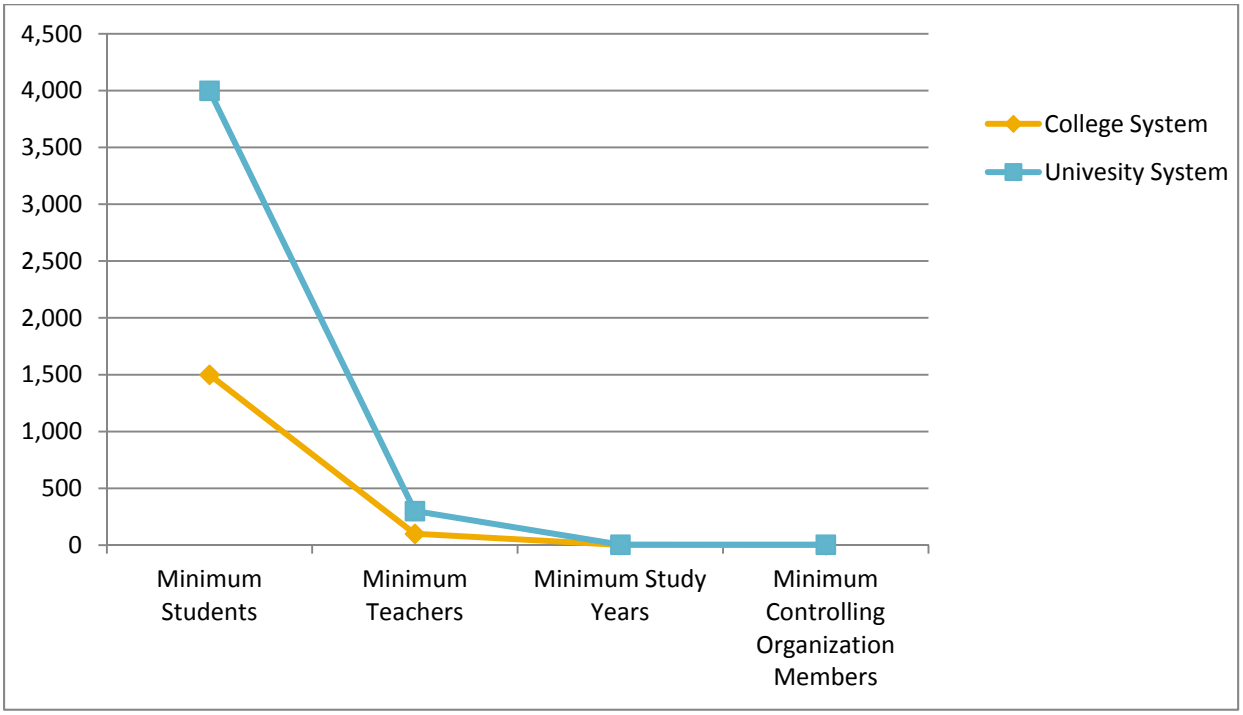
IT is used as business partner in University System. Basically, we analyzed University secure the student future more as compare to College system. Investor of the University will get more benefits as compare to College System. So success rate in University system is more than College system. Better decision makes the system successful. Our main purpose is to achieve the business value successfully, so we did through proper analysis of both systems (College and University) in education sector.

### 3.2 Graphically

We statistically analyzed the College and University differences. We analyzed generally between College and University System through numerical data.

We focused on four components such as:

- Minimum Students
- Minimum Teachers
- Minimum Study Years
- Minimum Controlling Organization Members




Graph 1

Therefore, University System is better than College System.

## 4. Conclusion

Information Systems are being progressively more used to value businesses. Developed nations are facing greater competitive stress because of well-organized and effective production which is at present standard in many sectors. We concluded that business value play most important role in Information Systems, so IS (Information System) can be applied in any area of business.

Well, Helpful Business Value of Advance BAL Information System (HBVABIS) is enhanced solution for business value of information system.

## References

[1]
http://en.wikipedia.org/wiki/Information_system

[2]
http://bisom.uncc.edu/courses/info2130/Topics/istypes.htm

[3]
The Price of Security: The Challenge of Measuring Business ValueInvestments in Securing Information Systems by: *Tony Coulson, Jake Zhu, Shan Miyuan, C.E. Tapie Rohm*
www.iima.org/CIIMA/8%205.4_Coulson_19_24.pdf

[4]
*Awais, M; Samin T; and Bilal, M.(2011).*Effective Business Value of Bal Information System. IJCSI 8( 6).Page(366-370)

# Covering Space and Van Kampen theory methods of Fundamental Group

Gbenga Olayinka Ojo[1#], Ajuebishi Patient Adetunji[1+] , Oluwaseun Gbenga Fadare[2*] ,
Fisayo Caleb Sangogboye[2†] , Hezekiah Oluleye Babatunde[3] and Funmilayo Ruth Abodunrin[2‡]

[1]Department of Mathematics and Statistics , Joseph Ayo Babalola University,
Ikeji-Arakeji, Osun-State, Nigeria.

[2]Department of Computer Science , Joseph Ayo Babalola University,
Ikeji-Arakeji, Osun-State, Nigeria.

[3]Department of Computer Science , Osun-State University,
Oshogbo, Osun-State, Nigeria.

**Abstract**

This paper examines two methods of computing fundamental group which are covering space method and the Van Kampen theory method. Van Kampen theory method is more analytical than the cover space method; the idea is used to solve a geometrical problem of global nature by first reducing it to homotopy theory problem which in turn reduces to an algebraic problem and solves as such. In 2001, H. Fausk et al [1] and Hu [2] showed the isomorphism between left and right adjoint theory and its application to homotopy categories, bearing in mind that homotopy group are higher dimension of fundamental group. In this paper detail explanations of fundamental group and homotopy group will be given. Various definitions and explanation of concepts, which are directly or indirectly related, will also be considered with illustration on how fundamental group can be calculated. This paper also reviews that in principle any space that can be broken up into pieces can have its fundamental group described by generators and relations via Van Kampen's theorem

*Keywords:* ***Homotopy Group, Fundamental Group, Van Kampen theory, Covering space, Isomorphism***

## 1. Introduction

The theory of fundamental groups and covering spaces is one of the few parts of algebraic topology that has probably reached definitive form, and it has not be generally treated in many sources. This review paper will explore its rudiments in actual sense.

In algebraic topology, homotopy theory is the study of homotopy groups, more generally of the category of topological spaces and homotopy classes of continuous mapping at an intuitive level, a homotopy class is a connected component of a function space, while homotopy group is said to be a higher dimension of fundamental group. Fundamental group is denoted as $\pi(X, x)$ which consists of all equivalence classes of loops based at $x$ and the product operation between them.

May J.P [3,4] defined topological space X as a set in which there is a notion of nearness of points, given a collection of open subsets of X which is closed under finite intersections and arbitrary unions. It is then suffice to imagine that metric spaces connotes open sets that are the arbitrary unions of finite intersections of neighbourhoods

$U\varepsilon(x) = \{y|d(x, y) < \varepsilon\}$.
A function p: X $\longrightarrow$ Y is continuous if it takes nearby points to nearby points, $p^{-1}(U)$ is open if U is open. If X and Y are metric spaces, this means that, for any $x \in X$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $p(U\delta(x)) \subset U\varepsilon(p(x))$. Algebraic topology assigns discrete algebraic invariants to topological spaces and continuous maps.

## 1.1 General topology

Let $\mathcal{X}$ be a non-empty set. A class $\tau$ of subset of $\mathcal{X}$ is a topology on $\mathcal{X}$ (point topology) iff $\tau$ satisfied the following axioms:
[01] X and $\emptyset$ belong to $\tau$ i.e., X, $\emptyset \in \tau$.
[02] The arbitrary union of any number of sets in $\tau$ belongs to $\tau$.
[03] The finite intersection of any two sets in $\tau$ belongs to $\tau$.
Therefore the pair $(X,\tau)$ is called a topological space.

## 1.2 Category

A category C consists of the following:
(a) A class of objects (family of set)
(b) For every ordered pair of objects A and B, a set Mov (A, B) of "Morphisms"
Iff $f \in$ Mov (A, B) we write: F: A $\rightarrow$ B or A $\xrightarrow{f}$ B
(c) For every ordered triple of objects A, B and C, a function is associated to a pair of morphisms f: A $\rightarrow$ B and g: B $\rightarrow$ C their "composite"
$g \circ f$ : A $\rightarrow$ C i.e. $f \in$ Mov(A,B), $g \in$ Mov(B,C) then $g \circ f \in$ Mov(A,C).

## 1.3 Covariant Functor

Let C and D be categories respectively, a covariant functor is a map $\tau$: C$\rightarrow$ D consisting of object function which assigns to every object $A \in$ C, an object T (A) $\in$ D and morphism function which assigns to every
$f \in$ Mov (A, B) $\in$ C a morphism T (f) $\in$ Mov (T(A), T(B)) Such that:
1. T (1A) = 1T (A) identity goes to identity
2. T ($g \circ f$) = T (g) · T (f) composite goes to composite.

## 1.4 Contravariant functor

If C and D be categories respectively, a Contravariant functor is a map $\delta$ : C $\rightarrow$ D consist of an object which assign to every object $A \in$ C an object S(A) $\in$ D and a morphism function which assigns to every
$f \in$ Mov (A, B) $\in$ C a morphism S(f) $\in$ Mov(S(B), S(A)) in D Such that:
1. S (1A) = 1S (A) identity goes to identity
2. S ($g \circ f$) = S (f) · S (g) composites goes to composite.

## 1.5 Exact sequence

An exact sequence consists of family Aq, $q \in Z$ of algebraic structure together with morphisms fq : Aq $\rightarrow$ Aq+1 such that we have a long sequence:
$\cdots Aq-1 \xrightarrow{fq-1} Aq \xrightarrow{fq} Aq+1 \xrightarrow{fq+1} Aq+2 \xrightarrow{fq+2} \cdots$
which is exact at every point of the sequence i.e. Im fq = kerfq+1 $\forall$ q.

## 1.6 Homotopy

Let X, Y $\in$ C, denoted by $Y^* = \{f: X \rightarrow Y / f \text{ is a map}\}$ which is continuous f, $g \in Y^*$ are said to be homotopic if there exist a map
f: X × I $\rightarrow$ Y, I = [0, 1] the unit interval such that:
F ($\mathcal{X}$, 0) = f ($\mathcal{X}$) $\forall \mathcal{X} \in$ X
F ($\mathcal{X}$, 1) = g ($\mathcal{X}$) $\forall \mathcal{X} \in$ X;
then F is said to be homotopy written as f $\sim$ g (f is homotopic to g).

## 1.7 Definition of Arc or Path

A path in a topological space X is a continuous map of some closed interval into X i.e. $f \in X^i$ = f: I $\rightarrow$ X such that: $x_0$ = f(0) to $x_1$ = f(1). $x_0$ and $x_1$ are initial and terminal point respectively.

Let X be a space and X1 andX2 are paths in X respectively such that
X1(1) = X2(0). Then the composite path X1 · X2 is given by

$$X1 \cdot X2 = \begin{cases} X_1(2s) & \text{if } 0 \le s \le \frac{1}{2} \\ X_2(2s-1) & \text{if } \frac{1}{2} \le s \le 1 \end{cases}$$

A space X is called arcwise connected or Pathwise connected if any two Points of X can be joined by an arc.

The path components of X are the maximal arcwise connected subsets of X (i.e. ordinary components of X).

If the map f: I $\rightarrow$ X is a path such that f (0) = f (1) is called a loop which is based at a point $\mathcal{X} \in$ X.

## 1.8 Retract and Deformation Retract

A is said to be a retract of X if the identity map of A, 1A can be extended
to map r: X $\rightarrow$ A i.e. A $\xrightarrow{i}$X $\xrightarrow{r}$ A.

Let A ⊂ B ⊂ X, A is called a deformation retract of B over X if 1B ∼ a
Retract r: B → A.

### 1.9 Covering Spaces

Let X be a topological space, a covering space is a space $\tilde{X}$ and a map
$\pi : \tilde{X} \rightarrow$ X such that:
(1) $\pi$ Is onto
(2) $\forall\ x \in$ X $\exists$ a neighbourhood V of □: $\pi^{-1}$(V) is disjoint union of Open
Sets each f which is mapped homeomorphically onto V by $\pi$ where X is the base space and $\tilde{X}$ is the total space.

Examples of covering space are:
(1) $\pi_n : S^1 \rightarrow S^1$ given by $\pi_n(Z) = Z^n$ covering space of n fold covering.
(2) $: R^n \rightarrow S^1 \times \cdots \times S^1$ given by $\pi$(X1. . . Xn) = $(e^{2\pi ix} \ldots e^{2\pi ixn})$
(3) $: [0, 1] \times R^1 \rightarrow [0, 1] \times S^1$
$\pi(s, t) = (s, e^{2\pi it})$
$[0, 1] \times S^{-1}$ Identity with $\{(x, y) \in R2 | 1 \leq x^2 + y^2 \leq 4\}$
If: $\tilde{X} \rightarrow$ X is a covering space such that $\tilde{X}$ is simply connected then the Covering space is called a universal covering space.

A connected space X is a space which is pathwise connected and whose fundamental group is trivial i.e. $\pi_1(X) = 0$.

### 1.10 Fundamental group

The class of map (homotopy class of map) $\pi(X, x_0)$ is referred to as the fundamental group for $x \in$ X. It is the set of all the loops based at $x_0$ which
For a group $\pi(X, x_0) = [I, (X, x_0)]$ {0, 1}.
If P: I → X then P (0) = P (1) it is a loop.

## 2.0 Calculating Fundamental Groups

Lewis et al [5] and May J.P [4] have supported that fundamental group can be analyzed and calculated by using two approaches:

### 2.1 Conversion of fundamental group problems to algebraic problems and solve as such, it is often achieved by putting algebraic structure on sets of homotopy.

Considering two methods of calculating fundamental group, the first method which is the covering spaces method is quite geometric and connections between the spaces is not necessary because it allow working based on intuition to the answer. The second method is the Van Kampen theorem which is analytical and some what used to show that the space is the map e: $H^1 \rightarrow S^1$ given by e (t) = $e^{2\pi it}$ e is periodic of period 1.We think of a spiral connected of this space into a circle.

For the calculation of fundamental group there is need to relate it with the structure of covering spaces with the path lifting property.

Path lifting property
Given P: I → X and a $\in \tilde{X}$ such that:
$\pi$(a) = P (0).
There is a unique Path $\tilde{P}$ : I → $\tilde{S}$ such that $\pi\tilde{P}$ = P and P (0) = a.

Example using covering space
Homomorphism $Z \rightarrow \pi_1$ ( $S^1, 1$ ): n ↦ $\alpha^n$ is an isomorphism the formula n ↦ $\alpha^n$ determines a homomorphism $Z \rightarrow \pi_1$ ($S^1, 1$ ) then show that loop s : I → $S^1$ starting at 1 is homotopic to $S_n$ if the path $\tilde{S}$ : I → $\mathbb{R}$ covering s and starting at 0 ∈ $\mathbb{R}$ end s at x ∈ $\mathbb{R}$.
Also that $S_n$ is null homotopic if n = 0.

### Solution

The map $Z \rightarrow \pi_1$ ( $S^1, 1$ ) is well defined homomorphism, by map s: I → $S^1$ it is an epimorphism and by map $\tilde{S}$ : I → $\mathbb{R}$ it is a monomorphism, therefore it is an isomorphism.
If n ↦ $\alpha^n$ and K ↦ $\alpha^K$, then n + k ↦ $\alpha^{n+K} = \alpha^n . \alpha^K$
$\mathbb{R}$ is connected, the paths $\tilde{S}$ and $\tilde{S}^n$ are homotopic therefore the the path S and $S_n$ are homotopic. [S] = [$S_n$] =$\alpha^n$. But if n ≠ 0 then the path is not a loop and loop $S_n$ is not null – homotopic.

This method is not analytical enough; therefore, we describe a tool for calculating $\pi(X, x_0)$.
Assume X = X1 ∪ X2 and X1 ∩ X2 ≠ ∅ then choosing $x_0$ ∈ X1 ∩ X2
We have:

$i_1 : \pi(X1 \cap X2, x_0) \rightarrow \pi(X1, x_0)$ and
$i_2 : \pi(X1 \cap X2, x_0) \rightarrow \pi(X2, x_0)$ which is homorphisms, making a general group construction.
Let G1 and G2 be groups f1 : G → G1 and f2 : G → G2 homorphism. The amalgamation of G1 and G2 over G is the smallest group generated by G1 and G2 with f1 ($x$) = f2 ($x$) for $x \in$ G.
If F is the free group generated by G1 ∪ G2 then:
$x \cdot y$ is three products in F
F is of the form $x_1^{\epsilon_1} \ldots x_k^{\epsilon_n}, \epsilon_1 = \pm 1$ and X ∈ G1 ∪ G2.

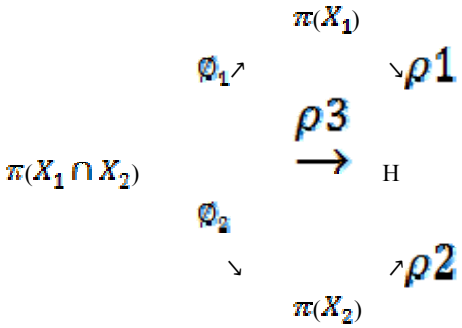
The Van Kampen theorem allows the calculation of $\pi(X, x_0)$ provided $\pi(X1)$, $\pi(X2)$ and $\pi(X1 \cap X2)$ are known.

### 2.1 Van Kampen Theory

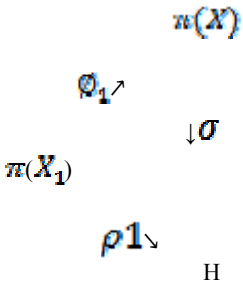
The statement and prove of the theorem Van Kampen theorem are as follows:
As X1 and X2 are connected space open subsets of X such that X = X1 ∪ X2 and X1 ∩ X2 = ∅ and are connected, choosing a base points $x_0 \in$ X1 ∪ X2 for all $\pi(X, x_0)$ under consideration.

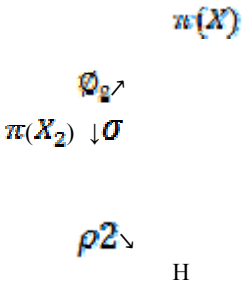
Theorem 2.1.1: Let H be any group and P1, P2, P3 are the homomorphism

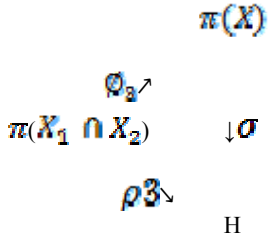
Such that:




There exists a unique homomorphism σ1 π(X) → H such that the diagrams is commutative



The prove of this theorem is to show that $\pi$(X) is a characterized up to Isomorphism by the theorem i.e
$\sigma : \pi(X) \rightarrow H$

Proof:
To achieve this we show that it is
Homorphism, it is one-on-one and onto.
Assume $x_1; x_2 \in H$.
Then $\sigma : \pi(X) \rightarrow H$ is well defined,
then $\sigma((x_1)(x_2)) = \sigma(x_1) : \sigma(x_2)$ by definition
$\sigma((x_1)(x_2)) = \sigma((x_1, x_2))$
$= (\alpha^{-1}x_1, x_2\alpha)$
$= ((\alpha^{-1}x_1\alpha))((\alpha^{-1}x_2\alpha))$
$= \sigma(x_1) \cdot \sigma(x_2)$
Therefore, it is homorphism.

Assume $\sigma(x_1) = \sigma(x_2)$
If $x_1 = x_2$ then it is 1 – 1.
Given $(\alpha^{-1}x_1\alpha) = (\alpha^{-1}x_2\alpha)$
$\Rightarrow \alpha^{-1}x_1\alpha \simeq \alpha^{-1}x_2\alpha$
$\Rightarrow x_1 = x_2$ hence it is 1 − 1

Let $x_2 \in H$ to determine $x_1 \in \pi(x)$ such that $\sigma(x_1) = x_2$.
Consider $x_1 = \alpha x_2 \alpha^{-1}$ a loop at $x_0$. Then
$\sigma(x_1) = \sigma(\alpha x_2 \alpha^{-1})$
$= (\alpha^{-1}(\alpha x_2 \alpha^{-1})\alpha)$
$= 1 \cdot x_2 \cdot 1$
$= x_2$.
Hence it is isomorphism.

The most general version of Van Kampen theorem consist of covering Space X by any number of open sets which is not just two open sets. This Open set must be arcwise connected; also the intersection of any finite number must be arcwise connected containing the base point.

Illustration of Van Kampen method
Let X be a space X = A [B, A \ B = $\{x_0\}$ and A and B are each Homeomorphic to circle S0, X may be visualized as follows
Let X1 = A1, X2 = B to determine the
Structure $\pi(x)$ but A and B is not open.
Let a $\in$ A and b $\in$ B: a $\neq x_0$ and b $\neq x_0$.
Let X1 = X − {b} and X2 = X − {a}, X1 and X2 are homeomorphic to a
Circle X1 \ X2 = X − {a, b} is contractible.
Hence they are simply connected.

Thus $\pi(x)$ is a free product of the group $\pi(x_1)$ and $\pi(x_2)$.
Thus $\pi$(A) and $\pi$(B) are infinite cyclic group.

## 2.2 Interpolation of Fundamental Group and Covering space

Suppose Z is a space, and * a point of Z. We define $\pi_1$(Z,*) as homotopy classes of maps f:[0,1] --> Z, such that f(0) = f(1) = *.

Shmuel [6] proved that the boundary conditions are absolutely critical for getting a nontrivial theory. $\pi_1$ (Z,*) is a group using concatenation of paths; the constant path is the identity and "going backwards is the inverse. $\pi_1$ (Z,*) is referred to as the fundamental group of Z. (If Z is path connected, the choice of * is irrelevant.

Example: If Z is the circle $S^1$ = {u in **C** | |u| = 1}, we can define a map $\pi_1$ ($S^1$,1) -> **Z** (the integers) by sending a map f to
(log(f)(1) – log(f)(0))/2 $\pi$i.

Definition. A map p: A -> B is a covering space, if:around each point b in B, there is a neighborhood N of b, so that $p^{-1}$ (N) is a disjoint union of sets Ai each of which is mapped homeomorphically onto N by p.
The map exp: **R** -> $S^1$ considered before is a good example.

Examples: The 2-sphere $S^2$ is simply connected. The projective plane $\mathbf{RP}^2$ has fundamental group **Z**/2**Z** since it is the quotient of $S^2$ by making the identifications x = -x. The projection map is a covering map, and the group of covering transformations is just **Z**/2**Z** = {id, x -> -x}. The nontrivial element in the fundamental group of **RP2** can be thought of
as the quotient of a great chord on S2 that connects the north pole to the south pole.

### 2.3 Computation of Fundamental Group.

Fundamental group discussed earlier has two popularilities; the first being its connection to covering space theory. The second is that it is quite computable that is Van Kampen theory
Example: If X is contractible then the fundamental group is trivial.
Example: If one sees the universal cover and group of deck transformations, then one also knows the fundamental group.

The practical tool of this computation is Van Kampen's theorem.

Van Kampen's theorem. Let Z denote the union of A and B, and X denote their intersection. If A, B, and X are all connected (and nonempty), and then $\pi_1$ (Z,x) is generated by $\pi_1$ (A,x) and $\pi_1$ (B,x). The only relations among the elements of $\pi_1$ (A,x) and $\pi_1$ (B,x) are the ones forced by the fact that the elements of $\pi_1$ (C,x) can be thought of as elements of both of these groups.

Examples.
1. If A and B are simply connected, and their intersection is connected, then their union is simply connected.
2. If X is simply connected, then $\pi_1$ (Z,x) is the free product
$\pi_1$ (A,x)* $\pi_1$ (B,x). The elements of the free products are just finite strings of elements of $\pi_1$ (A,x) and $\pi_1$ (B,x), and one multiplies strings by concatenting them, ignoring the identity, and combining contiguous elements of the same group.

3. These groups can be tricky if $\pi_1$ (X,x) is nontrivial. The group described is called a free product with amalgamation and is denoted by $\pi_1$(A,x)* $\pi_1$ (X,x) $\pi_1$ (B,x).

Interpretation of this is that the elements of this look like when the induced maps of $\pi_1$ (X,x) into the other two pieces are injective, but without this it can get complicated. As a simple example suppose that X is a circle and that $\pi_1$ (A,x) = **Z**/2**Z** and $\pi_1$ (B,x) = **Z**/3**Z,** so that the induced homomorphisms are the obvious surjections. Hence, space is established and that Van Kampen's theorem tells us that Z is simply connected.

The fundamental groups of both A and B are generated by that of the circle, i.e. there is one generator, say g. From A we learn that $g^2 = e$ and from B we learn that $g^3 = e$. So in the amalgamated free product (i.e. $\pi_1$ (Z,x)) g = e, so the who group vanishes.

## Conclusion

Fundamental group have been treated geometrically, it was formulated in a simple way with algebraic convention and some of its concepts and theorems such as: concept relating homotopy maps with homotopy group were briefly reviewed. The higher dimension of this fundamental group is applicable to spaces such as: Real projective spaces, Complex projective spaces, Moore space M ($\pi$, n) etc., but for this paper calculation of fundamental group was only discussed because this is necessary before the application. In principle any space that can be broken up into pieces can have its fundamental group described by generators and relations via Van Kampen's theorem and then calculated appropriately.

References

[1] H. Fausk, P. Hu, and J.P. May. Isomorphisms between left and right adjoints. Preprint, 2001.

[2] P. Hu. Duality for smooth families in equivariant stable homotopy theory. Preprint, 2001.

[3] J.P. May. The additivity of traces in triangulated categories. Advances in Mathematics 163(2001), 34-73.

[4] J.P May. " A concise course in Algebraic Topology available at http://www.math.uchicago.edu/~may/CONCISE/conciseRevised.pdf

[5] L. G. Lewis, Jr., J. P. May, and M. Steinberger (with contributions by J. E. McClure). Equivariant stable homotopy theory. Lecture Notes in Mathematics Vol. 1213. Springer. 1986.

[6] Shmuel. Fundamental group and convering Space: the facts. Available at: http://www.math.uchicago.edu/~shmuel/fund.pdf

# Cloud computing and its applications in the world of networking

**Puja Dhar**[1]

[1]**Department of Information Technology,I.T.S-Management & IT Institute, Ghaziabad,Uttar Pradesh 201007,India**

**Abstract**

The paper discusses the most discussed topic nowadays 'cloud computing'. Cloud computing is becoming a buzzword. The paper explains the concept, Services provided by cloud computing and different service providers. Also it works out how this technology can be harnessed to bring benefits to the challenges of the business, in terms of cost reduction and maintain competitiveness.

***Keywords:*** *Grid computing, IaaS, PaaS, SaaS,*

## 1. Introduction

The term 'cloud computing' refers to computing services available to anyone online. Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.

## 2. Components of Cloud computing

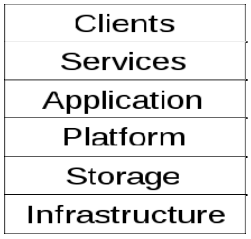



Fig 1: Components of Cloud Computing

### 2.1 Clients

A cloud client consists of computer hardware and/or computer software which relies on cloud computing for application delivery, or which is specifically designed for delivery of cloud services and which, in either case, is essentially useless without it.

### 2.2 Services

A cloud service includes "products, services and solutions that are delivered and consumed in real-time over the Internet. For example, Web Services which may be accessed by other cloud computing components and software.

### 2.3 Applications

A cloud application leverages the Cloud in software architecture, often eliminating the need to install and run the application on the customer's own computer, thus alleviating the burden of software maintenance, ongoing operation, and support.

### 2.4 Platform

A cloud platform, such as Platform as a service, the delivery of a computing platform, and/or solution stack as a service, facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software.

### 2.5 Storage

Cloud storage involves the delivery of data storage as a service, including database-like services, often billed on a utility computing basis.

### 2.6 Infrastructure

Cloud infrastructure, such as Infrastructure as a service, is the delivery of computer infrastructure, typically a platform virtualization environment, as a service

## 3. Services Provided by Cloud Computing

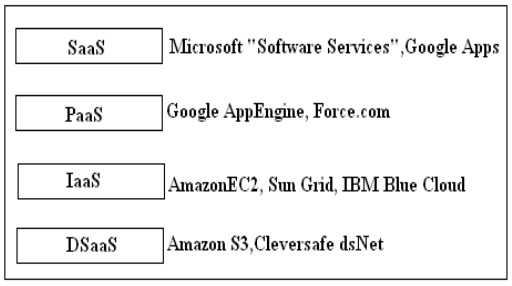



Fig 2: Services of Cloud computing

### 3.1Software as a Service (SaaS)

The software applications like CRM, Office Suite, Email, etc., are offered as a service through the internet, instead of a shrink wrapped software on a physical medium (or in a downloadable form), which is the norm in the traditional desktop world. The applications are hosted on a highly scalable infrastructure and it is offered over the internet. Users can access it using an ordinary web browser, without any need to install software in their local computer. Companies like Google, Zoho, Salesforce, Microsoft, Wordpress offer their applications as a service to the end users.

### 3.2 Platform as a Service (PaaS)

Some vendors are offering application development platform as a service. Developers can code the applications and upload it into the platform (offered as a service) and run the application on the cloud infrastructure. It helps developers to scale their apps without worrying about building the infrastructure. The platform scales automatically based on the resource needs of the app, without any efforts from the developer. Services like Google App Engine, Bungee Connect and Force.com are examples for PaaS.

### 3.3 Infrastructure/Hardware as a Service (HaaS)

Vendors offer computing infrastructure as a service to end users. The term Hardware as a Service is a bit of a misnomer. It is actually computing power offered through a virtualized environment rather than a physical hardware. This service is offered either as raw computing power or storage or both. Some examples of services offered in this category include Amazon's EC2 and S3, Mozy, GoGrid, etc.

## 4. Cloud Computing Vs. Grid Computing

Grid computing is the application of several computers to a single problem at the same time - usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data.

Grid computing depends on software to divide and divide up pieces of a program among several computers, sometimes up to many thousands

This technology has been applied to computationally intensive scientific, mathematical, and academic problems through volunteer computing, and it is used in commercial enterprises for such diverse applications as drug discovery, economic forecasting, seismic analysis, and back-office data processing in support of e-commerce and Web services.

Cloud computing is Internet ("cloud") based development and use of computer technology ("computing"). It is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.

Users need not have knowledge of, expertise in, or control over the technology infrastructure "in the cloud" that supports them.

## 5. Working of Cloud Computing

In theory the process is very simple. Cloud computing could allow you to have only a small computer, inexpensive computer, processor and monitor in front of you. You would have no need for a hard drive or a CD/DVD drive. Instead you would need only an Internet connection, which would hook you up to a central supercomputer that would host all your programs and files. This presents and advantage to both storage and security issues.

## 6. Example of Cloud Computing Usage

There are different examples of cloud computing like Email communication now plays a central role in most of our busy lives. We carry a mobile WiFi-enabled laptop with us everywhere we go or use push email on our cell phone, having an email client sitting on our computer at home means that while out and about us risk spending time outside of the communication loop. This is one area where the cloud finds it's most frequent and useful application.

Sometimes you may find yourself in need of the opinion of your colleagues. Downloading files onto flash memory, emailing documents to friends or family or colleagues or sending submissions by mail. Google launched a service that allowed groups of people to work on the same document, idea or proposal in real time or whenever convenient to each participant.

Using Google Wave you can create a document and then invite others to comment, amend, offer opinion, or otherwise join in with the creation of the final draft. Google is not alone in producing online collaboration tools. Other examples include Spicebird, Mikogo, Stixy and Vyew to name but a few.

For the dedicated cloud enthusiast, something like Amazon's EC2 virtual computing environment might be the answer to all our needs. Rather than purchasing servers, software, network equipment and so on, users would buy into a fully outsourced set of online services instead.

Most cloud environments on offer can customize the kind of service provided to exactly suit the needs of the user. If we need more processing power from time to time, a cloud-based infrastructure, being scalable, negates the need for up-front investment in client-owned resources.

Data stored on our home or business computer suffers from many of the same restrictions as email and, as with email, the cloud offers a solution. Storing our MP3's, video, photos and documents online instead of at home gives use the freedom to access them wherever we can find the means to get online.

### 6.1 Examples of online storage services

It includes *Humyo, ZumoDrive, Microsoft's SkyDrive, S3* from Amazon, amongst others. Many offer both free and paid for storage and backup solutions.

### 6.2 Google as fore runner of cloud computing

Online suite of office applications is probably the best known but by no means the only solution on offer. Rather than having a system and space hogging suite of applications like a word processor, a spreadsheet creator and a presentation or publishing platform sitting on your computer, you could opt to work online instead. Accessibility, potential for collaboration and perhaps even online storage are just some of the benefits of satisfying your office suite needs by working online.

Google is currently the fore runner of cloud computing due to its need to produce accurate and instant results for the millions of search queries it receives every day.

## 7. Properties of Cloud Computing

There are different key properties of cloud computing like

- ✓ Task-Centric
- ✓ Powerful
- ✓ User-Centric
- ✓ Programmable
- ✓ Accessible

## 8. Major Service Providers of Cloud Computing

### 8.1Google 101-Network

Made up of millions of cheap servers, that would store staggering amounts of data, including numerous copies of worldwide web. It makes search faster, helping ferret out answers to millions of queries in a fraction of a second.

### 8.2 Microsoft's Azure

It is a Internet-scale cloud computing and services platform hosted in MS data centers. It provides a range of functionality to build applications that span from consumer web to enterprise scenarios.

### 8.3 Amazon's Elastic Compute Cloud-Amazon EC2

This is a web service interface that provides resizable computing capacity in a cloud. It is designed to make web-scale computing easy for developers. It allows developers to pay only for capacity that they actually use.

### 8.4 IBM's CloudBrust

It is developed for the everyday user. IBM also offers private cloud computing services using IBM blue services software

## 9. Applications and advantages of Cloud Computing

There are various applications of cloud computing in today's network world. Many search engines and social websites are using the concept of cloud computing like www.amazon.com, hotmail.com, facebook.com, linkedln.com etc. the advantages of cloud computing in context to scalability is like reduced risk , low cost testing ,ability to segment the customer base and auto-scaling based on application load.

## 10. India to use the 'Cloud' Services for e-governance

India is to become one of the first countries in the world to deliver e-Governance services to citizens using cloud-based IT services. The government is in talks with software industry body, Nasscom, on the roll-out of e-Governance services using the emerging technology. The advantage of using this technology is that the IT infrastructure need not be set up by the government. In addition, because of the ability of the technology to handle large number of transactions, citizens can look forward to less congestion bottlenecks.

## 11. Conclusions

Cloud computing will promote the use of shared resources and when we are sharing the resources among different users it will definitely lower the costs and will help in keeping the environment clean. Cloud computing will also help in e-learning by providing many services online for the students. We need to strap up this technology in our daily lives by creating many applications on cloud.

## References

[1] Anita Campbell (2008-08-31). "Cloud Computing-Get Used to the Term" The App Gap. http://www.theappgap.com
[2] Williams John M. Chris sears (2008-21-31). "Who Coined the Phrase Cloud Computing?".
[3]"Web Services Glossary" http://www.w3.org/TR/ws-gloss.
[4] http://developeramozonwebservices.com
[5] www.egovonline.net
[6] How to Secure Cloud Computing. http://searchsecurity.techtarget.com/magOnline/0,sid14_gci1349551.html.

**First Author** She is having experience of more than six years in academics. She is working as Assistant Professor in Department of Information Technology at I.T.S Mohan Nagar, She has published number of papers in proceedings of national and international conferences and refereed journals. She has done M.Tech(IT),Msc(IT). She has received 'Best Faculty Award' at I.T.S. Her research interest includes networking and databases.

# A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment

V.THAVAVEL and S.SIVAKUMAR*

Department of Computer Applications, Karunya University, Coimbatore, Tamilnadu, India-641 114.

*Corresponding Author

**Abstract**

The management of unstructured data is recognized as one of the major unsolved problems in the information industry and data mining paradigm. Unstructured data in computerized information that either does not have a data model and there are not easily usable by data mining. This paper proposes a solution to this problem by managing unstructured data in to structured data using legacy system and distributed data partitioned method for gives distributed data for mining multi text documents. This frame work gives the testing of the similarities among text documents and privacy preserving meta data hiding technique, which are explored in text mining.

***Keywords:*** *Unstructured data, Privacy preserving data mining, Distributed data mining, Testing Similarity.*

## 1 Introduction

Privacy preserving distributed data mining is the extraction of relevant knowledge from large amount of data, while protecting at the same time sensitive information or personally identifiable information in the unstructured distributed data environment. Terrovitis[14] have adapted group-based methods such as k-anonymity to "unstructured" data by treating text data as a sort of variable length database record, or set of un-typed values, with the assumption that the sensitive value to protect is deterministically contained in this set. Chris Clifton[3] address the problem of data is vertically partitioned and privacy means preventing others from learning the value of private attribute values for each entity. The need for privacy preservation is privacy of source because unauthorized user interact to damage the data of misuse of information and to support heterogeneity of source. Clifton[15] motivate vertically data separation techniques for distributed environment and working with structured data not a unstructured data environment. This paper aim to develop a privacy preserving distributed data mining frame work for unstructured data environment and to achieve a device of new structured data model from NETMARK. It helps data integration for unstructured data; protect Meta data by use hiding technique. This paper attempt to distribute heterogeneous data to the network by using vertically data separation method and it enables text mining to test the similarity measure among text documents.

This paper has proposed and designed a new generalized frame work for privacy preservation in distributed data mining for unstructured data environment and implementing the testing of similarity among free form of text using testing of hypothesis (Inferential Statistics). In this paper, section 2 describes the literature work, section 3 provides designing of a new frame work for privacy preservation in distributed data mining for unstructured data environment, section 4 discusses implementation of a frame work and testing the similarity of the given documents and finally, section 5 gives the conclusion of the work.

## 2 Literature Review

### 2.1 Privacy preserving data mining

A number of approaches and techniques such as randomization, and k-anonymity have been developed in order to carry out privacy-preserving data mining task.

#### 2.1.1 Randomization method

In randomization method for privacy-preserving data mining, the noise is added to the data set in order to mask the attribute values of sensitive record fields[1][2]. The amount of noise added is large enough to smear original values, so individual records values cannot be recovered techniques are designed to derive aggregate distributions from the perturbed records. Subsequently data mining methods can be developed in order to work with these aggregate distributions. The key advantage of the randomization method lies in its simplicity as method does not require knowledge of the distribution of other records in the data, unlike other methods such as k-anonymity which require the knowledge of other records in the data. Therefore, the randomization can be applied at data collection time without the use of a trusted server containing all original records.

Kargupta et.al[7] challenges perturbation and randomization –based approaches. They claim that such approaches may lose information as well as not provide privacy by introducing random noise to the data by using random matrix properties, Kargupta et al. successfully separates the data from the random noise and subsequently discloses the original data.

#### 2.1.2 The k- Anonymity model

The K-anonymity model [12] was proposed to deal with the possibility of indirect identification of records form public databases. Since combinations of records attributes can be used to exactly identify individual records. In k-anonymity the granularity of data representation is reduced by employing techniques such as generalization and suppression. The granularity is reduced to such a level that any given record maps onto a least K other records in the dataset. The k-anonymity method was first proposed by Samarati [11]. The approach uses domain generalization hierarchies of the quasi-identifiers in order to build K-anonymous tables. The concept of K-Minimal generalization has been proposed Samarati [11] in order to limit to level of possible for a given level of anonymity.

#### 2.1.3 Secure multi-party computation

Secure multi-party computation (SMC) deals with general problem of functions secure computation with distributed inputs. In privacy preserving data mining the solutions that posses the rigor of work in SMC settings and typically make use of cryptographic techniques are known as SMC solutions.

Yao first formulated the two-party comparison problem (Yao's Millionaire protocol) and presented a provably secure solution [13]. It was extended to multiparty computations by Goldreich et al.[6]. They developed a frame work for secure multiparty computations, and in [5] proved that computing a function privately is equivalent to computing if securely. A semi honest party(also known as honest but curious) follows the rules of the protocol using its correct input, but is free to later use what it sees during execution of the protocol to compromise security.

### 2.2 privacy preserving Distributed Data mining

The primary role of distributed methods for privacy preserving data mining is to enable computation of useful aggregate statistics over the joint databases without compromising the privacy of the individual datasets within the different participants. So, the participants may wish to collaborate in obtaining aggregate results, may not fully trust each other in terms of the distribution of their own databases Lindell and Pinkas[7] first introduced this technique to the data mining community. Their method enabled two parties to jointly contract a decision tree without either party gaining any knowledge about each other's data except what might be revealed through the final decision tree. Specifically they targeted ID3 Algorithm with horizontally partitioned data. For the purpose of privacy, the datasets may be partitioned either horizontally or vertically. In case of horizontally partitioned datasets, the individual records are spread out across multiple entities, each have the same set of attributes. In vertical partitioning the individual entities may have different attributes (or view) of the same record sets. Chris Clifton[3] deals with finding to address the problem of association rule discovery, where data is vertically partitioned, and privacy means preventing others from learning the value of private attribute values for each entity. The problem of distributed privacy preserving data mining closely resembles field of cryptography for determining secure multi-party computations and share common techniques [10].

### 2.3 Unstructured Data Environment

Miller[8] have designed a system to facilitate the interaction of structured and unstructured data. The main features of the view mechanism, especially as they relate to textual documents are presented in the paper. It also looks at how the views approach allows the interaction between the data taken from structured (example: Relational) semi structured (object oriented) and unstructured (example: Text) data source. Data mining and how the system will operate in the complete environment. This paper, describe in extensible view system that support integration of both data from heterogeneous structured and unstructured data sources in either the multi database or data ware house environment. First Approach is makes use of a global schema; it typically makes use of common data model and a global languages. Second Approach is multi database language approach common language to define how the data sources are integrated, transferred and presented. Third Approach is the local site works closely with a set of inter-related sites to setup the partial global schema.

David A.Maluf [4] deal with NETMARK was NASA Ames Research centre designed and developed a data management and integration system. NET MARK that achieves data integration

across multiple structured and unstructured data source in a highly scalable and cost efficient manner.

Querying and integration of originally unstructured data such as various formatted report in micro soft word, Adobe portable Document Format(PDF), Excel Spread sheets and power point presentations, is a key focus , given that the bulk of enterprise data is indeed un structured

## 3. Proposed Method

The objective of this paper is proposed a generalize framework for the privacy preservation distributed data mining for unstructured environment.

### 3.1 Unstructured data into Structured Data Environment

It deal with converting the unstructured in to structured data, the unstructured data is converted to Xml, node representation and relational storage with Meta data.

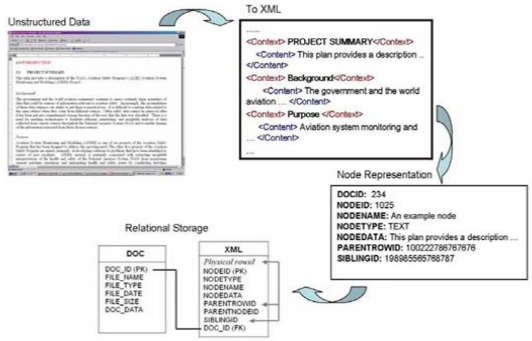



**Figure 1 Unstructured data into Structured Data**

### 3.2 Distributed Mechanism

Distributed mechanism was designed for real storage of metadata with text. Distribute data as method of vertically separation techniques for heterogeneous data.

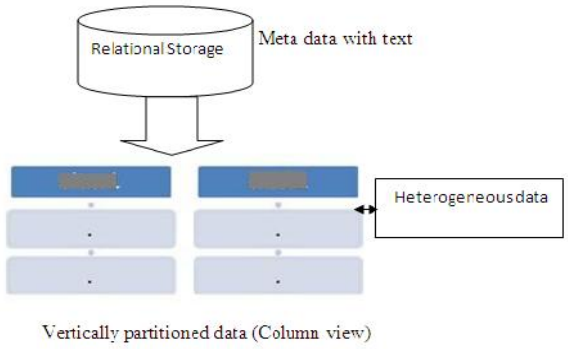



**Figure 2 Distributed Mechanisms**

### 3.3 Security Mechanism

It deals with to apply in Meta data with hiding techniques for privacy policy.

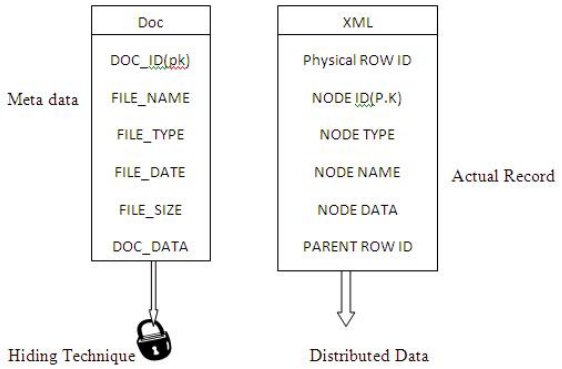



**Figure 3 Security Mechanism with hiding Meta data for privacy preservation**

### 3.4 Text Mining

It perform text processing to find words or attributes in documents occurrences in word list process document from files

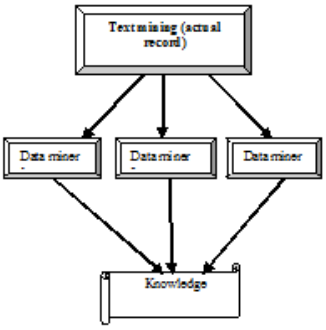



**Figure 4** Text process is distributed to data miner for Text mining

### 3.5. Privacy Preservation in Distributed Data mining for Unstructured Data Environment

Design and develop a data model (structured data) from unstructured data, managing unstructured data are converted into XML and then create data table contain Meta data.

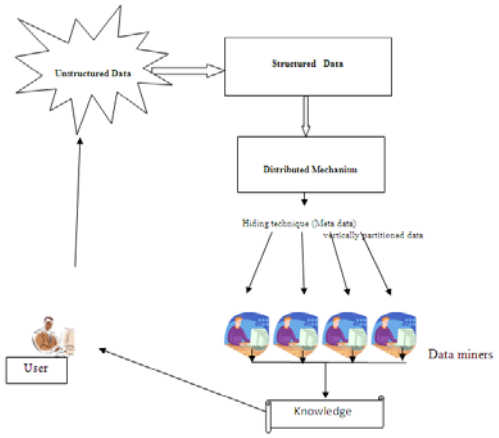



**Figure 5 Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment**

Meta data like data about data in textual information, For example file name, Date of creation, file type, file size, Author of document etc., Because of this conversion is used to well efficient of distributed data mining of textual area and secure of personally identifiable information. Data table contains Meta data and xml data from various text files is stored in relational Storage. Relational storage data (heterogeneous) is distributed along network path with more securely. Privacy preserving distributed data mining distribute as datasets like horizontally or vertically partitioned. In this frame work distribute only xml data (Real storage data with column view) by using vertically partitioned method. Privacy policy is applied to Meta data with information hiding technique for security purpose. Data miner only knows xml text information not a personally identifiable information and also intruder does not interact with personal data without authentication. Xml data is distributed to data miners; data miners perform information extraction and measure text document similarity. In this frame work using text mining is extract knowledge from heterogeneous data into related data groups. (i.e. here clustering method is extract similar/related data group from heterogeneous data).Distributed data mining paradigm have more one than data miner to perform knowledge extraction process, at the same time interaction among one miner to another for sharing data mining intermediate results not a source data. In this frame work perform automatic discovery of new, previously unknown, information from unstructured textual data. Finally develop a generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment.

## 4. IMPLEMENTATION

Unstructured data environment converted structured data environment and privacy preservation of Meta- data designed by using VB.NET application. The implementation setup considered the text documents (.txt file extension) only. In this paper

take two text file documents both size as 42.4kb (43,497 bytes) and 19.7kb (20,178 bytes) was used to produce the resultant shown below. Actual text processing as like document occurrences and total occurrences of words in text document was designed by using Rapid miner tool. Example dataset (Two examples, 4 special attributes, 2199 regular attributes), where two examples are m1.txt and m2.txt.Special attributes are Labels (me9, me10 are type is binominal), Metadata-file, metadata-path (type is polynomial) and Meta data-date (type is date-time). Remaining are regular attributes of both text files in words or attribute names. The following screenshot shows the working environment of this frame work

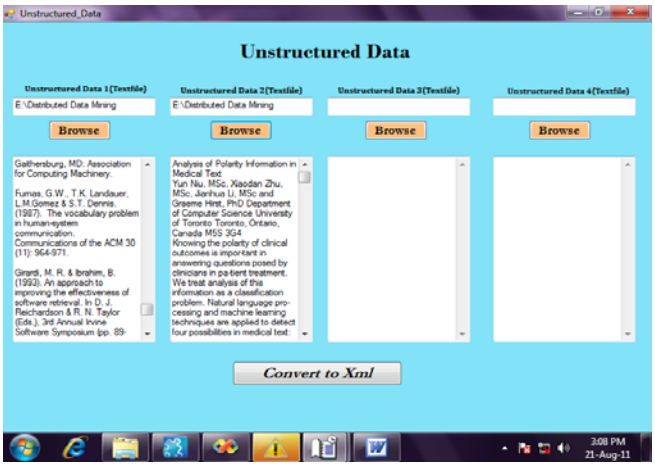



**Figure.6.Unstructured data (Text file) browse from node and apply operation convert to XML and Meta data format**

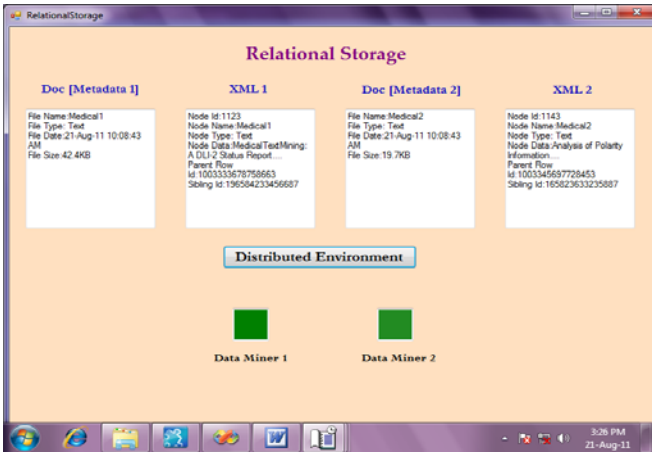



**Figure.7. XML are distribute through distributed environment into data miners instant of hiding Meta data**

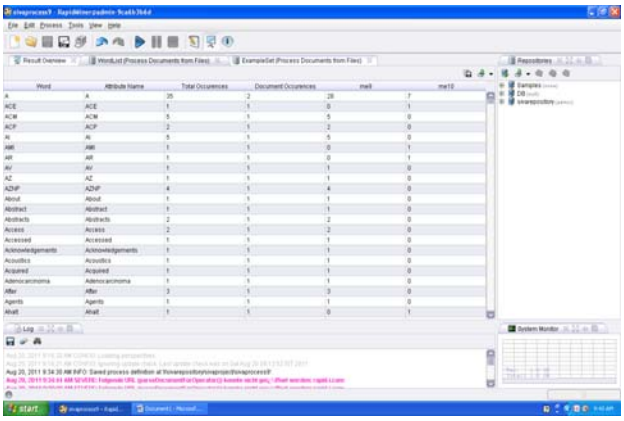


**Figure.8. Text data processed by text processing using rapid miner tool find words or attributes in documents occurrences in word list process documents from files**

### 4.1 Result and Discussion

For testing the similarity of given documents using the method of testing of hypothesis (Inferential Statistics). This paper deals the testing of two documents, namely m1 and m2. Words occurrences of the above said documents are 2199 regular attributes.

$$Z_c = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where $Z_c$ is the calculated value of standard normal variate. $\bar{x}$ and $\bar{y}$ are the mean values of the words occurrences in documents m1 and m2 respectively. $\sigma_1^2$ and $\sigma_2^2$ are the variances of the words occurrences in documents m1 and m2 respectively. $n_1$ and $n_2$ are the number of regular attributes of documents m1 and m2 respectively. $Z_t$ = 1.96 is the table value of standard normal variate at 5% level of significance. Let us assume that, there is no similarity in documents m1 and m2 (null hypothesis). Here $Z_c$ = 2.936789 , $\bar{y}$ = 1.398818, $\sigma_1^2$ =163.3763, $\sigma_2^2$ =38.69847 and $n_1=n_2$= 2199 and $Z_c$ = 5.0735, therefore $Z_c > Z_t$, so rejects null hypothesis. Hence the documents m1 and m2 are Similar.

**4. Conclusion and Future work**

` This paper provides frame work for privacy preservation of Meta data using hiding technique in unstructured data environment with a distributed mechanism. The proposed system is also perform text processing to find words or attributes occurrences in documents and testing similarity measure using normal distribution method which was discussed. In future generalized frame work is extend to manage and analyzing privacy preservation for distributed data mining in unstructured data like e-mail messages, complicated reports, presentations, voice mail, still images, and video.

**5 References:**

[1] Agrawal. D and Aggarwal .C.C, *"On the Design and Quantification of Privacy Preserving Data Mining Algorithms":ACM PODS Conference*,(2002).

[2] Agrawal.R and Srikant.R, *"Privacy-Preserving Data Mining";ACM SIGMOD Conferene,pp*.439-450,Dallas,TX,May 14-19,(2000).

[3] Chris Clifton, *Privacy Preserving Distributed Data mining,* Department of Computer Science, Nov 9,2001*www.cs.purdue.edu/homes/**clifton**/.../**Clifton**DDM.pdf*

[4] David A.Maluf and Peter B.Tran, *Managing Unstructured Data with Structured Legacy Systems*, 2008 IEEE.

[5] Goldreich.O, *The Foundations of Cryptography*, volume 2,chapter General Crytographic Protocols. Cambridge University Press,2004.

[6] Goldreich.O, Micali.S and A.Wigderson, *How to play any mental game- a completeness theorem for protocols with honest majority*, In 19th ACM Symposium on the Theory of Computing, pages 218-229,1987.

[7].Kargupta.H, Datta.S. Q.Wang and K.Sivakumar, *"On the privacy preserving properties of random data perturbation techniques"*,IEEE ICDM,2003.

[8].Lindell.Y and B.Pinkas. Privacy preserving data mining. Journal of Cryptology,15(3):177-206,2002.

[9].Park.B and H.Kargupta, "Distributed data mining": Algorithm,System,and applications.N.Ye,editor,The Hand book of Data mining,Pages 341-358.Lawrence Erlbaum Associates, Mahwah,N.J.,(2003).

[10] Pinkas. B, *Cryptographic Techniques for Privacy-Preserving Data Mining*,ACM SIGKDD Explorations,4(2),2002.

[11] Samarati.P, *"Protecting Respondents Identities in Microdata Release*, IEEE Trans.Knowl.Data Eng.13(6):1010-1027(2001).

[12] Samarati.P and Sweeney.L,"*Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppression"*, IEEE Symposium on Security and Privacy,(1998).

[13] Yao.A.C, *"How to generate and exchange secrets"*, In Proceedings of the 27th IEEE Symposium on Foundations of Computer Science, Pages 162-167.IEEE, 1986.

[14] Terrovitis. M, Mamoulis. N and Kalnis. P, *"Privacy-Preserving Anonymization of Set-Valued Data". In Proc. of VLDB Endowment,* 1(1), 2008.

[15] Chris Clifton, Murat kartarcioglu, Jaideep vaidya,Xiaodong Lin and Michael Y.Zhu, *Tools for Privacy Preserving Distributed Data mining*, SIGKDD Eplor. Vol 4,Issue 2,2002.

# High Performance Charge Pump Phase-Locked Loop with Low Current Mismatch

Prof.V.Sujatha,[1] Professor & Head,Dept.of ECE, Shree Sathyam Colege of Engineering and Technology,Sankari,Salem(Dt),Tamilnadu.
Dr.R.S.D.Wahida Banu[2], Professor&Head,Dept.of ECE, Govt.College of Engineering, Salem Tamilnadu.

**Abstract**

In CMOS CPs, which have Up and Down switches made of p-channel and n-channel respectively, generates fluctuations in the VCO due to current mismatch occurs when dumping the charge to the loop filter and subsequently a large phase noise on the PLL output. This paper presents a new CP circuit after detailed analysis of the current mismatch problem. It combines an error amplifier with reference current sources to achieve good current matching characteristics and lower phase noises. Charge sharing can be eliminated by using charge removal transistors. In addition, a low-voltage cascode current mirror and gain-boosting circuit are used to enhance current matching over process corners and increase the output impedance of the CP. Good current matching characteristic is achieved with less than 0.1% difference of the Up/Down current and 1% over all process variations. The CP output compliance voltage range of 0.1-1.8 V is achieved for 1.8-V supply voltage. The circuit was designed using 0.18um TSMC CMOS technology and simulated by Spectre tools.

***Keywords:*** *Charge pump (CP), gain-boosting charge pump, Voltage Controlled Oscillator (VCO) low-voltage cascode current mirror, phase-locked loop (PLL).*

## 1. Introduction

Since the conception of phase locking was proposed in the Thirties of the 20th Century, it has been widely applied in electronics and communication fields[1], especially used in large scale digital circuits. CP-PLLs (Charge Pump Phase-Locked Loop) are mainly used to generate signals and renew the clock pulses during the data transmission with high speed [5] [6]. As a key model, the charge pump plays an important role in assuring PLLs stability. It converts the digital signals in PFDs (Phase Frequency Detector) into analog signals of VCOs (Voltage Controlled Oscillator). When the phase-locked loop was locked in a certain frequency, the output voltage of charge pump is demanded to be a fixed value, and any tiny change of which will result in apparent frequency offset. Therefore, it is very important to design a charge pump circuit which can send a stable output voltage in CP-PLLs plan.

Phase-locked loops are widely used in clock generators and RF transceivers to ensure the accuracy of the oscillator frequency. The charge pump (CP) is an essential block in phase-locked loops (PLL). The CP consists of two switched current sources. Any current mismatch between the two current sources (i.e., difference between the source and the sink currents) would cause ripples on the control voltage. Ripples result in large phase noise and would also cause spurs on the PLL output signal [2].

This paper proposes a modified technique to decrease current mismatch in CP's using gain-boosting in addition to low-voltage cascode current mirrors [1]. Section II reviews the basic idea of the CP and the reasons behind current mismatch. Section III and IV deals with Charge sharing problem and current matching characteristics respectively. Section V introduces the proposed circuit architecture. Finally simulation results and comparisons in section VI. Section VII concludes the paper.

## 2. Basic Principle

Fig. 1 shows the circuit diagram of a conventional charge pump. In the charge pump, the digital output signals (UP and DN) of the PFD control the two circuit sources (IUP and IDN), and charge the capacitance CL via two switches which generally substituted by two MOSFET to obtain the DC level Vctrl needed by the Voltage Controlled Oscillator.

In Fig.1 the IUP and IDN should be completely equal in theory, but there are many nonideal effects which will result in their

mismatching in practice [4]. Another ubiquitous problem is the charge sharing in conventional charge pumps which results from the parasitic capacitance of node A and B. The level in node A will be charged to VDD, and in node B be discharged to GND when the signal UP and DN are invalid, whereas the node A level will be falling and the node B level will be rising when the signal UP and DN are valid. The difference between Vctrl and node A will not be uniform to the difference between Vctrl and node B, thus bring on the charge redistribution among CL, A and B. Because the Ids will change with Vds, current source IUP or IDN will share the charge. It will result in current mismatch which make Vctrl jittering, and influence the circuit performance. The output Vctrl would be held if the charge and discharge current are well matched. Generally, the net current generated by the charge pump is not equal to zero because of the current mismatching, it will make the Vctrl increase a fixed value in every phase comparing time. The control voltage Vctrl should be held in an average value to maintain the loop in a locked status (as shown in Fig. 2(a)), then the phase-locked loop shall bring on phase error which make the net current of the charge pump be zero in every period, as shown in Fig. 2(b) where A1 and A2 have the equal area.

The phase error resulted from the current mismatch can be expressed as the following formula where Δton, Icp and |IUP- IDN| respectively represent the dead zone time of the PFD, the period of the reference clock, and the offset between CP current and charge, discharge current.

The eq.(1) indicates that, to lower the phase error, the dead zone time Δton and mismatch current |IUP-IDN| should be reduced, but the CP current Icp should be increased while the reference clock period is fixed. Holding a definite dead zone time will be propitious to overcoming the PFD dead zone, and the higher current Icp will increase the power consumption and noise, so lessening the mismatch current |IUP-IDN| is the key to lower the CP phase error.

### 2.1. Charge Sharing Problem

For the charge pump in Fig. 3(a) (Type A), charge sharing is caused by the parasitic capacitance in nodes *pcs* and *ncs* [7]. When $I_{UP}$ is active, node *pcs* is charged to $V_{DD}$ When deactivating $I_{UP}$ some of the charge stored in node *pcs* will leak through the current source device. Since the parasitic of nodes *ncs* and *pcs* can never be matched, this will lead to a static phase offset. This is the transfer function of a phase-frequency detector followed by a Type A charge pump. The two transistors Mp and Mn in the Type B charge pump in Fig. 3(b) will remove the charge from the nodes *pcs* and *ncs* when Up and Down are deactivated [8].This leads to a large reduction in the phase offset.

Fig:2. Jitter transfer functions for different division ratios. (a) Simulated standard PLL. (b) Measured characteristics of Loop A with intentionally low damping.

Through the Mp device to the output when the Down control is inactive. When NMOS devices are used for speed-regulating the VCO, Vvco will never drop below Vtn, constraining Vqbn to be less than 2 Vtn which can easily be fulfilled. However, the charge pump works only up to an output voltage of Vvco< Vqbp + Vtp, limiting the upper tuning range of the VCO. However, the charge pump in Fig. 3(a) has the same upper voltage limit. Mismatch in $I_{UP}$ and $I_{DOWN}$ is a similar source of jitter as charge sharing described above. For low jitter, it is essential to have good matching, implying that the devices controlled by Vqbn/Vqbp should be saturated. Again, this requires Vvco< Vqbp + Vtp.

Charge removal can also be done by ac coupling [9], but this requires careful timing of the control signals in the charge pump. The solution to charge sharing in [10] is less suitable for low-applications due to the common-mode restrictions on the differential amplifier.

### 2.2. Current Matching Characteristics

Current matching characteristics is achieved by using an error amplifier as shown in Fig.4,, the voltage Vref, at the node REF of the current mirror *(M5* - Ms) follows the voltage *Vcpout* at the node CPOUT of the charge pump (M1~ *M4).* As a result, the voltage *Vref,* is equal to the voltage *Vcpout* as long as the amplifier maintains a high enough gain. For *M5* = M1, *M6* = *M2, M7* = M3 and M8 = *M4,* if the DOWN and the *UP* signal are high, then I4 = *I3* = *I2,* and if the DOWN and the UP' signal are low, then I3 = *I2* = I1. So we can make the sinking current *I4* equal the sourcing current Il. In this way, one can achieve nearly perfect source/sinking current matching characteristics regardless of the charge pump output voltages.

The proposed novel charge pump circuit which use error amplifier and reference current source to obtain the improved characteristic for

current match, and reduce the PLL's phase noise. Simultaneously, the charge sharing is effectively restrained by using charge removal transistors. So the circuit possesses good current match and high working speed. Current mirrors are then used to copy currents from the bias cell, and a cascode is used at the output node to get high output impedance in order to reduce current mismatch.

By using the gain-boosting technique, shown in Fig.5, high output impedance can be achieved without adding more cascode devices [1]. Gain-boosting saves some voltage headroom; this is significant for short channel length technologies, which have low supply voltage.

**2.3 Proposed Circuit**

The proposed CP circuit greatly improved the circuit performance by enhancing the current match and lessening the charge sharing, and at the same time, possessed the characteristics of high operating speed and low power consumption. In Fig.5, capacitance C1, C2 fill the role of stabilizing the node E and F's voltage to avoid instantaneous grid voltage fluctuation of the current source. The voltage Vc in node C will change with Vctrl by inserting an error amplifier which has the gain high enough to make Vc equal Vctrl. Moreover, M11 is designed to equal M9, M8 equal M10, M3 equal M5, and M4 equal M6, so the current I4 will be the same as I3, I1 when UP, DN level is holding high, and I2 will be the same as I1, I3 when UP, DN level is holding low.

Finally the current I4 equals I2, which lead up to the result of almost perfect drain-source current matching. M7 and M12 are named as charge eliminating transistor. When the transistors transfer from saturation to cutoff, the charge resting on the channel will be emitted to the source, and that the drain will not be impacted. When the UP and DN is low, the spare charge will be removed from node A and B, so that the charge sharing can be successfully restrained.

The proposed CP circuit which use error amplifier and reference current source to get good match characteristic, and use charge removal transistor to restrain the charge sharing. All the design has been simulated with Spectre tools.

The shortcoming of the structure is that it will confine the dynamic range, but it is not important in most situations. While Vctrl less than Vg5-Vtn and DN invalid, the current will move to the output via M7. The Vctrl will not be less than Vtn during NMOS used as VCO current control, and forcing the Vg5 less than 2Vtn will be easy to implement. Because the Vctrl is up to Vg10+|Vtp|, it restricts the Vctrl range. Transistor M7 and M12 also improve the switch speed of current, and supply DC level for node A, B while switch is turned off, which will prevent the pending nodes influencing the control Voltage.

A modification for the gain-boosting CP is illustrated in Fig.5, in which the channel length modulation problem is solved as well as current mismatch. A low voltage cascode current mirror is used to copy *IUp* and *IDown* from a single current source to ensure that both currents are equal. Fig.5.shows the low-voltage cascode current mirror. This current mirror is chosen because it provides high output impedance, and low channel length modulation mismatch [1].

As shown in Fig.5, the UpB and Down controlled switches have been embedded in the low-voltage cascade current mirrors, where M7 and M8 are the UpB and Down controlled switches respectively, M9 to M16 are the lowvoltage cascode current mirror transistors, and finally M17 and M18 are the reference current generators. The bias voltage of the low-voltage cascode current mirrors is chosen to be GND for the PMOS switched mirror, and VDD for the NMOS switched mirror, as these are the values of UpB and Down signals at lock. This provides better matching.

The output impedance, *Rout*, is higher than that given in Eqn**(2)**, since there is another cascode transistor, *Rout* is given by

$$Rout = R_{06}.gm_5.R_{05} \text{ at the node D.} \quad \text{----- (2)}$$

This value is large, and so large length transistors are not needed for having high output impedance.

**2.4 Simulation Results**

The proposed charge pump circuit, at the 1.8V power source, is simulated with Spectre tools. Fig.6. shows the simulation result of CP charge circuit, and Fig.7 -11 shows the stability, gain of VCO output and the charge pump current with reference and feedback signal. The graph has been plotted for both conventional and proposed charge pump PLL. The maximum difference of the Up/Down current is less than 0.1%. The proposed modified gain-boosting CP is designed using 1.8V CMOS transistors, good current matching is observed in addition to a wide compliance voltage range of 0.1-1.7 V**,** which relaxes the VCO design. The maximum value for the current mismatch is less than 0.1%. Current mismatch result of proposed CP and gain boosting CP is shown in table 1. The proposed design shows the highest gain and more stability.

## 3. Tables, Figures and Equations

### 3.1 Tables and Figures

| Proposed Charge Pump | | Gain Boosting CP | |
|---|---|---|---|
| Ip(μA) | Mismatch() | Ip(μA) | Mismatch() |
| 16.23 | 0.61 | 14.5 | 0.8 |

**Table 1: Current mismatch results**

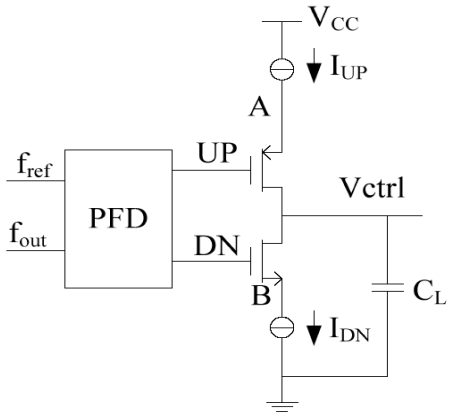



**Figure.1 conventional charge pump schematic**

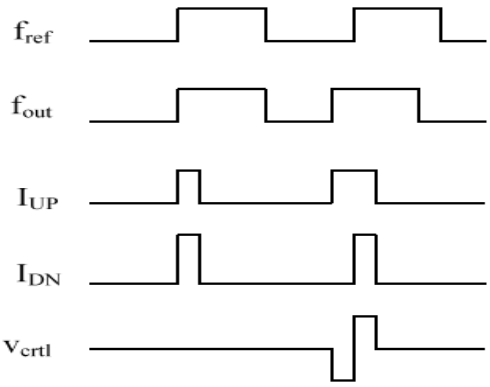



**Figure 2(a): Charge and discharge current mismatch**

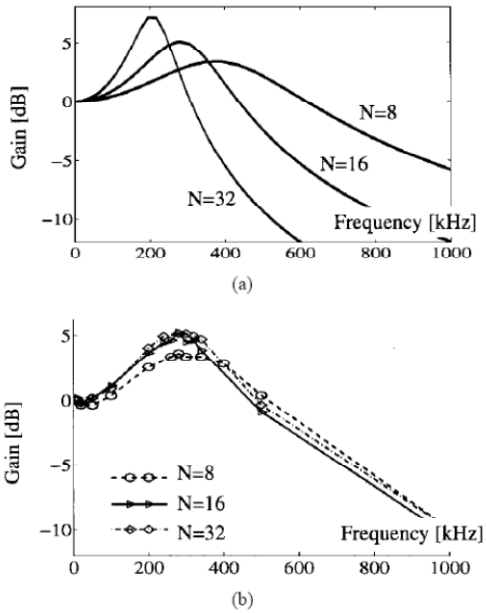



**Figure 2(b): Jitter transfer functions for different division ratios. a) Simulated standard PLL. (b) Measured characteristics of Loop A with intentionally low damping**

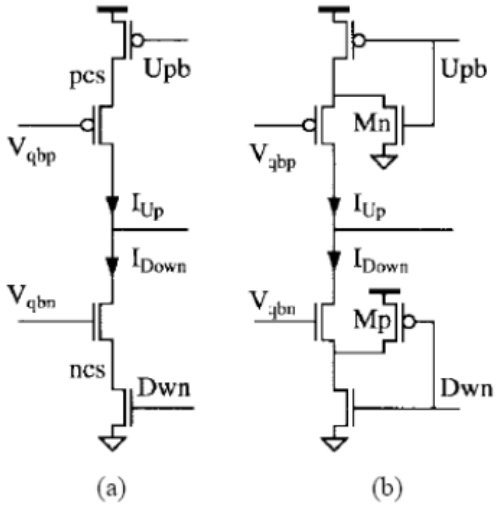



**Figure 3: (a) Charge-pump suffering from charge sharing (Type A). (b) Charge removal transistors eliminate charge sharing (Type B).**

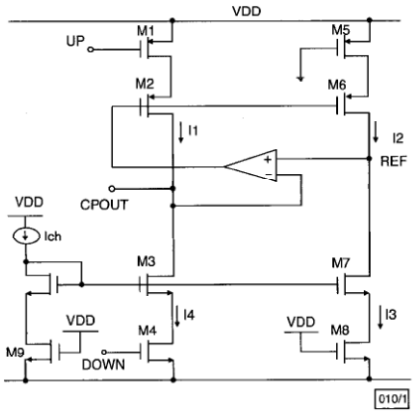



**Figure 4: Charge pump circuit with good current matching**

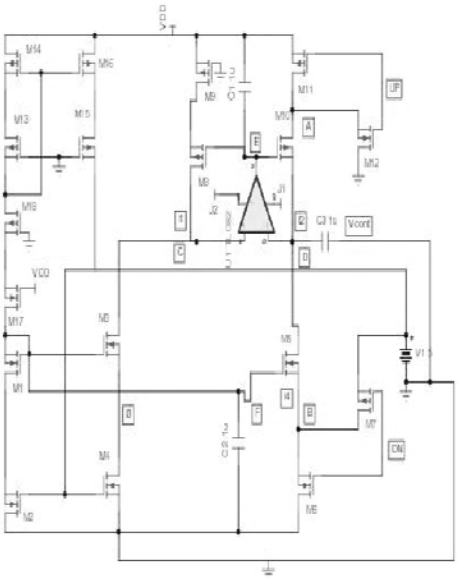


**Figure 5: Proposed CP circuit**

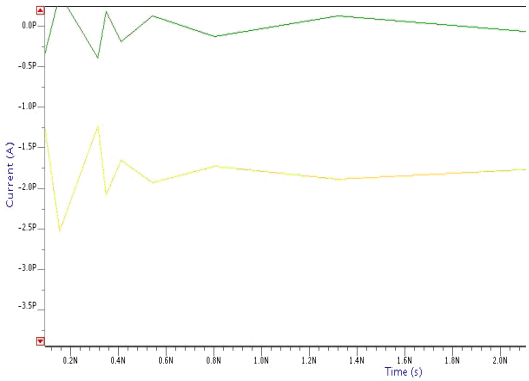



**Figure 6: IUp and IDown currents Plots**

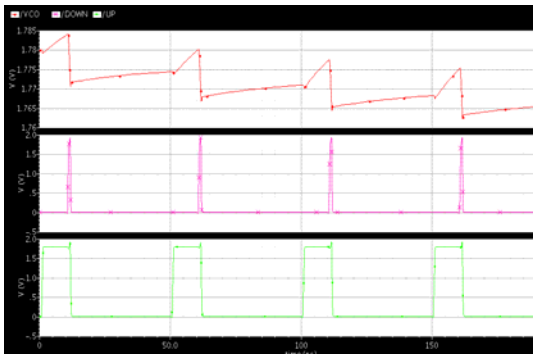


**Figure 7: VCO control voltage when reference signal leads feedback signal (Conventional Charge Pump PLL)**

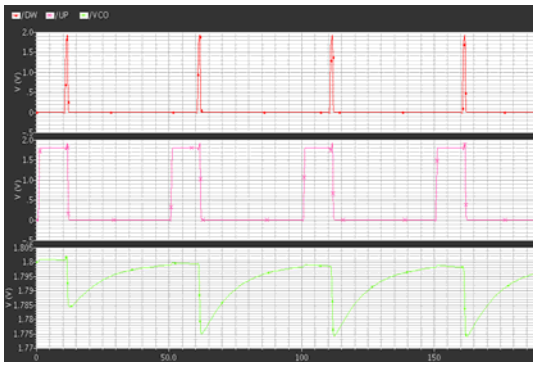


**Figure 8: VCO control voltage when reference signal leads feedback signal (Proposed Charge Pump PLL)**

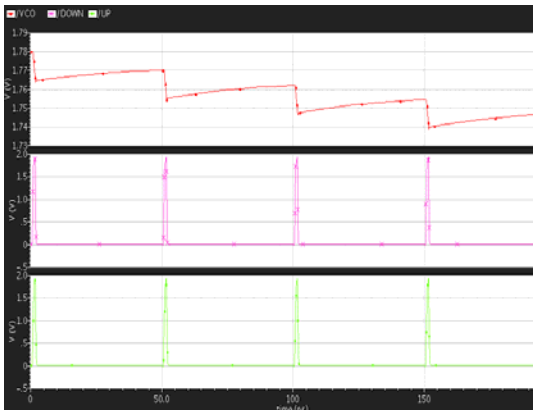


**Figure 9: VCO control voltage when reference signal and feedback signal phase & frequencies are equal (Conventional Charge Pump PLL)**

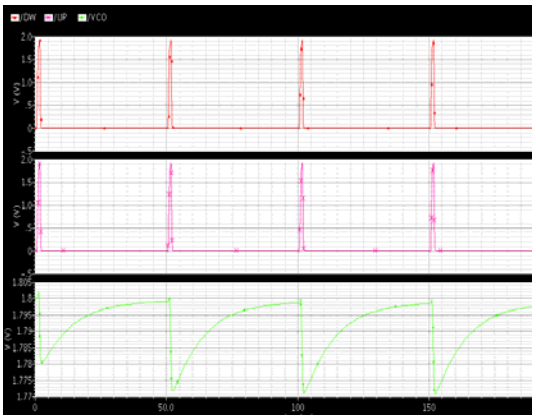


**Figure 10: VCO control voltage when reference signal and feedback signal phase & frequencies are equal (Proposed Charge Pump PLL)**

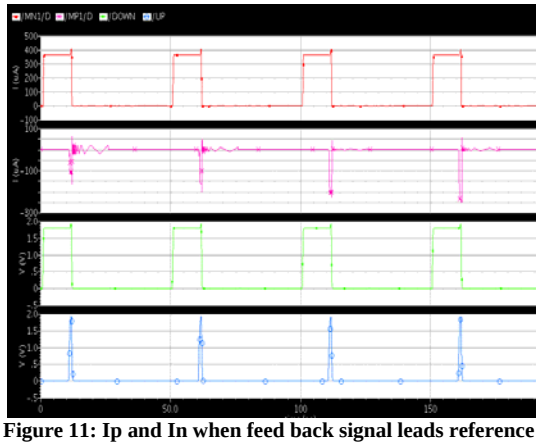


**Figure 11: Ip and In when feed back signal leads reference signal in the proposed CP.**

3.2 Equations

$$\Phi_{\varepsilon} = 2\pi \cdot \frac{\Delta t_{on}}{T_{ref}} \cdot \frac{\left| I_{UP} - I_{DN} \right|}{I_{cp}}$$

**Equation (1)**

$$\boldsymbol{Rout = R06.gm5.R05 \; at \; the \; node \; D}$$

**Equation (2)**

## 4. Conclusion

By using the gain-boosting and low-voltage cascode current mirrors, a high-performance low-mismatch charge pump is achieved. Good current matching characteristics can be achieved with less than 0.1% difference of the Up/Down current and 1% over all process variations. The CP output compliance voltage range of 0.1-1.8 V is achieved for 1.8-V supply voltage. Charge sharing problem can also be eliminated with the help of charge removal transistor and this proposed CP PLL also provides high gain by increasing the output impedance with the help of low voltage cascode current mirror. The circuit was designed using 0.18um TSMC CMOS technology and simulated by Spectre tools.

REFERENCES

1. Behzad Razavi, "Integrated Circuit Design of Analog CMOS, " XIAN, XIAN JIAOTONG University Publishing Company, 2002, pp. 432-470
2. W. Rhee, "Design of high-performance CMOS charge pumps in phase-locked loops," *IEEE Int. Symp. Circuits and Systems,* Vol. 1, 1999, pp. 545-548
3. P. Larsson, "A 2-1600 MHz CMOS Clock Recovery PLL with Low-Vdd Capability," *IEEE Journal of Solid-State Circuits*, Vol. 34, No. 12 (1999), pp. 1951-1959.
4. Kyung-Soo Ha and Lee-Sup Kim, "Charge-pump reducing current mismatch in DLLs and PLLs," *ISCAS 2006 IEEE International Symposium on Circuits and Systems,* May. 2006, pp. 21-24
5. Mark Van Paemel, "Analysis of a Charge-pump PLL: A New Model," *IEEE Journal of Solid-State Circuits,* Vol. 42, No. 7 (1994), pp. 2490-2498
6. Qu qiang, Zeng lieguang, "A dead zone Phase Frequency Detector used in Phase-locked Loop with high speed," *Micro Computer Information*, Vol. 12, No. 2 (2006), pp. 235-237
[7] M. Johnson and E. Hudson, "A variable delay line PLL for CPUcoprocessor synchronization," *IEEE J. Solid-State Circuits*, vol. SC-23, pp. 1218–1223, Oct. 1988.
[8] P. Larsson and J.-Y. Lee, "A 400 mW 50–380 MHz CMOS programmable clock recovery circuit," in *Proc. IEEE ASIC Conf. Exhibit*, 1995, pp. 271–274.
[9] V. von Kaenel, D. Aebisher, C. Piguet, and E. Dijkstra, "A 320 MHz, 1.5 mW at 1.35 V CMOS PLL for microprocessor clock generation," in *Proc.*

*IEEE Int. Solid-State Circuits Conf.*, 1996, pp. 132–133.
[10] M. Johnson and E. Hudson, "A variable delayline PLL for CPU coprocessor synchronization," *IEEE J. Solid-State Circuits*, vol. SC-23,pp. 1218–1223, Oct. 1988.

**V.Sujatha** received M.E degree in Applied Electronics from Anna University,Chennai in 2004 and the Bachelor degree in Electronics & Communication Engineering from Madras University in 1993.She is working as a Professor and Head of the Department of electronics and communication Engineering, Shree Sathyam College of Engineering and Technology,Sankari.She is a member of ISTE, IETE and IEEE. She published several papers in National and International conferences. Her area of research is low jitter charge pump PLL and testing of PLL.

**Dr. R.S.D. Wahida Banu, M.E.,Ph.D.**

Received the Ph.D. degree in Engineering from Anna University, Chennai, in 1998. Her areas of interest are Network security, Neural networks, system-on-a-chip design, She is a life member of IEEE,ISTE,CSI,SSI and member in ISOC and ,VDAT.She is a co-Author of the books titled as Object Oriented Programming Visual Programming, Data mining application for empowering Knowledge societies. To her credit she authored and co-authored not less than 200 research papers in International, National journals and Conferences. She bagged the Best Woman Engineer Award- 2009, Salem Local Chapter, Indian Institute of Engineers. Currently she is working as a Professor and Head of the department of Electronics and Communication Engineering, Government college of Engineering, Salem.

# A Novel Chaotic Encryption Scheme based on Pseudorandom Bit Padding

**Sodeif Ahadpour, Yaser Sadra and Zahra ArastehFard**

**Department of Sciences, University of Mohaghegh Ardabili, Ardabil, IRAN.**

**Abstract**

Cryptography is always very important in data origin authentications, entity authentication, data integrity and confidentiality. In recent years, a variety of chaotic cryptographic schemes have been proposed. These schemes have typical structure which performed the permutation and the diffusion stages, alternatively. The random number generators are intransitive in cryptographic schemes and be used in the diffusion functions of the image encryption for diffused pixels of plain image. In this paper, we propose a chaotic encryption scheme based on pseudorandom bit padding that the bits be generated by a novel logistic pseudorandom image algorithm. To evaluate the security of the cipher image of this scheme, the key space analysis, the correlation of two adjacent pixels and differential attack were performed. This scheme tries to improve the problem of failure of encryption such as small key space and level of security.

***Keywords:*** *Cryptography, chaos, Image Padding*

## 1. Introduction

The Cryptography is always very important in data origin authentications, entity authentication, data integrity and confidentiality [1-6,28,29,30]. In recent years, the cryptographic schemes have suggested some new and efficient ways to develop secure image encryption [1]. These schemes have typical structure which performed the permutation and the diffusion stages alternatively. However, most of algorithms be faced with some problems such as the lack of robustness and security. The random number generators are intransitive in cryptography for generation of cryptographic keys, allegorically, secret keys utilized in symmetric cryptosystems [2,3] and large numbers is intransitive in asymmetric cryptosystems [4,6], because of unpredictable, should better be generated randomly. In addition, random number generators in many cryptographic protocols, such as to create challenges, blinding value are used [7,8,9]. Also, the random number generators are used more in the diffusion functions of the image encryption for diffused pixels of plain image.

Random number generators can be classified into three classes which are pseudorandom number generators (PRNGs), true random number generators (TRNGs) and hybrid random number generators (HRNGs). PRNGs use deterministic processes to generate a series of outputs from an initial seed state [10,11,12]. TRNGs use of non-deterministic source (i.e., the entropy source), along with some processing function (i.e., the entropy distillation process) to generate the random bit sequence [2]. These sources consist of physical phenomena such as atmospheric noise, thermal noise, radioactive decay and even coin-tossing [13]. Many PRNGs using chaotic maps have been established. Most of them have very complex structures. In this paper, we propose a chaotic encryption scheme based on pseudorandom bit padding that the bits be generated by a novel logistic pseudorandom image algorithm. The random bit sequences produced by this generator are evaluated using the 15 statistical tests recommended by U.S. NIST [2]. Experimental results show that this PRNG possess good uniformity and randomness properties.

This paper is arranged as follows. In section 2, the properties of the logistic map are discussed. In section 3, we introduce the proposed random number generators and then discuss the uniformity and randomness of the bit sequences generated by the Proposed PRNG. In section 4, we propose chaotic encryption scheme based on pseudorandom bit padding and finally, in Section 5, we conclude the paper.

## 2. The logistic map

The logistic map is one of the most studied discrete chaotic maps. It is well-known as very sensitive to both system variable and control parameter. In addition, other features such as ergodicity, pseudo-randomness and

unpredictable behavior. Therefore, it possesses great potential for various cryptographic applications such as image encryption [15,16], public key cryptography [17], key agreement protocol [9], block cipher [3], and hash function [18,21,22]. It was first proposed as pseudo random number generator by Von Neumann in 1947 partly because it had a known algebraic distribution and mentioned later, in 1969, by Knuth [23,24]. The simplest form of the logistic map is given by:

$$x_{n+1} = rx_n(1 - x_n)$$

Where $x_n \in (0,1)$ and r are the system variable and control parameter, respectively, and n is the number of iterations. Thus, given a control parameter r and a system value $x_0$; time series of logistic map $\{x_n\}_{n=0}^{\infty}$ is computed. Here, we refer to $x_0$ and r as the initial state of the logistic map. In the following we use the chaotic logistic map for cryptographic applications, as follows:

$$x_{n+1} = rx_n(1 - x_n) \qquad (1)$$
$$x_n \in (0,1), \quad \text{and} \quad r_x \in (3.99996, 4]$$

[25]. As stated in [26],The choice of r in the equation above guarantees the existence of a chaotic orbit that can be shadowed by only one map as stated in . In addition, the above map is supposed to have good qualities as a PRNG when $r \cong 4$ [25].

## 3. The proposed PRNG and randomness analysis

### 3.1 The proposed PRNG using logistic pseudorandom image algorithm

Subheadings In this section, we introduce a proposed pseudorandom number generator based on the logistic pseudorandom image algorithm. For cryptographic purposes, the output of RNGs needs to be unpredictable [2]. In this method, we use a black white dynamic image because we can use each pixel as a key. On the other hand, key space of the PRNG is a black white dynamic image. To consider a gray scale image with the size of $2^k \times 2^k$ (here, $2^8 \times 2^8$) pixels (see Fig.1(a)).

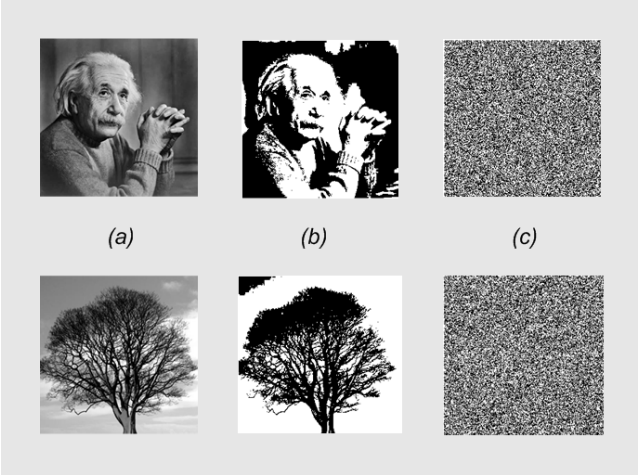



Fig. 1 (a) two gray scale images with the size of $2^8 \times 2^8$ pixels, (b) two black white images that the color of all pixels that are smaller than the Average Pixel Intensity (API) with black and all pixels that are greater than or equal to the API with white are changed, (c) two black white dynamic images that are the perfect seeds for PRNG.

We redefined it as a matrix $\mathbf{C}_{2^k \times 2^k}$. This matrix is composed of color of the pixels in the uint8 (output range 0 to 255). Uint8 is a MATLAB built-in function. Matrix components corresponding to image pixels can be showed as $\mathbf{c}_{ij}$. Then, we can get Average Pixel Intensity (API) [5,14]. Hence,

$$\mathbf{API} = \frac{\sum_{i=1}^{2^k}\sum_{j=1}^{2^k} \mathbf{c}_{ij}}{2^k \times 2^k}, \qquad (2)$$

then, we change the color of all pixels that are smaller than the Average Pixel Intensity(API) with black and all pixels that are greater than or equal to the API with white (see Fig.1(b)). Now, using a two-dimensional chaotic system which is defined as follows:

$$x_n = f(x_{n-1}) \qquad n = 0,1,2,...$$
$$y_n = g(y_{n-1}) \qquad n = 0,1,2,...$$

**that $f$ and $g : I \to I$ $(I = [0,1])$ are nonlinear maps, we get coordinates of a point $(x_n, y_n)$ in two-dimensional space. Using the following transformation can be converted coordinates of a point $(x_n, y_n)$ in two-dimensional continuous space into a point $(u(x_n), u(y_n))$ in two-dimensional discrete space of the image matrix components:**

$$u : I \to N \quad I = [0,1], \; N = [1, 2^k]$$
$$(i,j) = \begin{cases} u(x_n) = [x_n \times (2^k - 1)] + 1 \\ u(y_n) = [y_n \times (2^k - 1)] + 1 \end{cases} \qquad (3)$$

where symbol of [] is the round function. Then, we change color of the pixel $c_{ij}$ with coordinates of $(i = u(x_n), j = u(y_n))$ into the opposite color, i.e., if color of the pixel be white, it changes black and vice versa. In other words, if black and white colors be showed 0 and 1, respectively, those can be changed the following method,

$$c_{ij} = \begin{cases} 0 \to 1 \\ \text{or} \quad . \\ 1 \to 0 \end{cases} \qquad (4)$$

We iterate this method (Eq. 3,4) M times. Matrix that is created with this method, we show $C'_{2^k \times 2^k}$. The $M$ value is related to the two tests. So that, we create two bit sequence from the matrix $C'_{2^k \times 2^k}$. The first bit sequence to join the rows of the matrix is formed and the second bit sequence to join the columns of the matrix is formed. If two bit sequences to satisfy Monobit Test and Serial Test (see Appendix) separately, then, the M value is the correct value. Consequently, the resulting black white image (the black white image of the matrix $C'_{2^k \times 2^k}$) is the perfect seed for PRNG (see Fig.1(c)). For generating random bit sequence from this method, we are using a two-dimensional chaotic system which is defined as follows:

$$x'_n = f'(x_{n-1}) \qquad n = 0,1,2,...$$
$$y'_n = g'(y_{n-1}) \qquad n = 0,1,2,...$$

that $f'$ and $g' : I \to I$ $(I = [0,1])$ are nonlinear maps. Thus, using the transformation of Eq.3, the random bit sequence $\{z_n\}_{n=0}^{\infty}$ is defined as follows:

$$z_n = \begin{cases} 0 & c'_{i'j'} = 0 \\ 1 & c'_{i'j'} = 1 \end{cases} \qquad (5)$$

and because the black white image be a black white dynamic image, after each iteration of the Eq. 5 with this method adds the following term:

$$c'_{i'j'} = \begin{cases} 0 \to 1 \\ \text{or} \quad . \\ 1 \to 0 \end{cases} \qquad (6)$$

Therefore, we get a black white dynamic image as a seed for the proposed PRNG. As an example, we consider the logistic map (1) for the functions of $f$, $f'$, $g$ and $g'$ $(x_n, x'_n, y_n, y'_n \in (0,1)$, and $r_x, r_{x'}, r_y, r_{y'} \in (3.99996, 4])$ (see Fig.2).

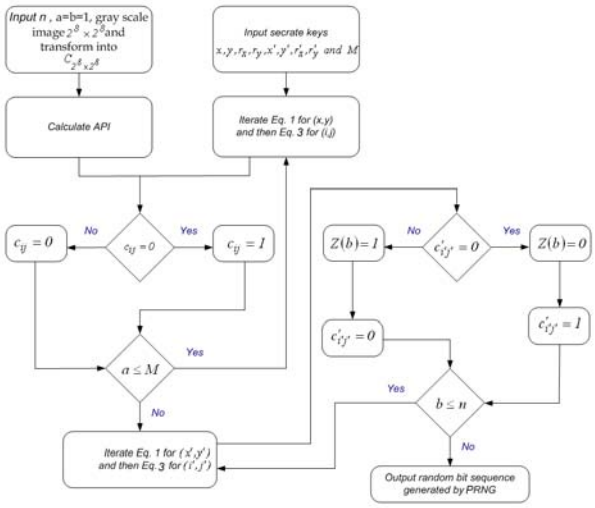



**Fig. 2. Block diagram of the logistic pseudorandom image algorithm for generation pseudorandom bit sequence.**

## 3.1 Analysis of randomness of number sequences

We have survey the randomness and uniformity of the several bit sequences of large size, generated by the proposed PRNG for different sets of control parameter and initial conditions of chaotic logistic maps and images. Here, we show the results for $2^{20}$ sized bit sequences corresponding to the following parameter values of the four sets:

$$\begin{cases} A = (0.2, 0.6, 4, 3.99997, 0.5, 0.1, 3.99998, 3.99999, 2^{18}) \\ B = (0.7, 0.3, 3.99998, 3.99996, 0.8, 0.4, 3.99997, 3.99999, 2^{18}) \\ C = (0.4, 0.8, 3.99996, 4, 0.6, 0.2, 3.99998, 3.99999, 2^{17}) \\ D = (0.7, 0.2, 3.99999, 3.99997, 0.3, 0.6, 3.99998, 4, 2^{17}) \end{cases}$$

For convenience, these four sets are designated as:

$$\{A,B,C,D = (x, y, r_x, r_y, x', y', r_{x'}, r_{y'}, M)$$

that A, B, C and D are related control parameter values of PRNG (see Table.1). We have used MATLAB 7.10.0 (R2010a) running program in a personal computer with a Core i3 3.1GHz intel, 4GB memory and 500GB hard-disk capacity. The average time used for generating random bit sequences with size of $2^{20}$ bits is shorter than 0.4 s.

We discuss in the following paragraph of this Section the result and conclusions of our study of the different statistical tests to observe the randomness and uniformity of the bit sequences generated by the proposed PRNG. The US NIST statistical test suite provides 15 statistical tests to detect deviations of a bit

Table I. Shows that the M value is the correct value if and only if two bit sequences created of the rows of the matrix $2^8 \times 2^8$ (1) and the columns of the matrix $2^8 \times 2^8$ (2) pass monobit test and serial test.

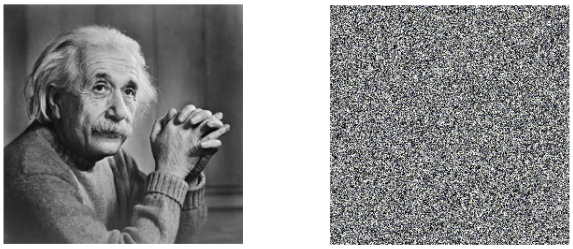
| Parameter | Calculated $\chi^2$ value | | | | Critical $\chi^2$ value at $\alpha = 0.05$ | |
|---|---|---|---|---|---|---|
| | Monobit test | | serial test | | Monobit test | serial test |
| | (1) | (2) | (1) | (2) | | |
| A (A. Einstein image) | 0.7985 | 1.0172 | 1.8143 | 1.8871 | 3.8415 | 5.9915 |
| B (Tree image) | 1.0378 | 1.1035 | 2.0568 | 2.0270 | 3.8415 | 5.9915 |
| C (A. Einstein image) | 1.0303 | 1.0098 | 2.1246 | 1.9883 | 3.8415 | 5.9915 |
| D (Tree image) | 1.4295 | 1.4487 | 2.4115 | 2.5635 | 3.8415 | 5.9915 |

Table II. Shows the P-values obtained from NIST suite for fifteen different tests. The P-values are obtained for four different sets of parameters for each test.

| NIST Tests | A (A. Einstein image) | B (Tree image) | C (A. Einstein image) | D (Tree image) |
|---|---|---|---|---|
| FT | 0.979743 | 0.600670 | 0.956387 | 0.284479 |
| FTB | 0.873583 | 0.794484 | 0.961466 | 0.218437 |
| RT | 0.863536 | 0.799775 | 0.766560 | 0.047121 |
| LROBT | 0.953186 | 0.643394 | 0.928064 | 0.287490 |
| BMRT | 0.920326 | 0.143269 | 0.273873 | 0.649518 |
| DFTT | 0.372087 | 0.544647 | 0.482314 | 0.214210 |
| NTMT | SUCCESS | SUCCESS | SUCCESS | SUCCESS |
| OTMT | 0.665345 | 0.093392 | 0.764690 | 0.399512 |
| MUST | 0.971350 | 0.165435 | 0.278815 | 0.218812 |
| LCT | 0.869026 | 0.424203 | 0.246919 | 0.597068 |
| ST P1 | 0.176425 | 0.807509 | 0.038659 | 0.155790 |
| P2 | 0.062528 | 0.867147 | 0.108128 | 0.355935 |
| AET | 0.198495 | 0.905032 | 0.548792 | 0.166571 |
| CST (FORWARD) | 0.999421 | 0.982586 | 0.977552 | 0.460996 |
| (REVERSE) | 0.998589 | 0.861198 | 0.991191 | 0.556137 |
| RET | SUCCESS | SUCCESS | SUCCESS | SUCCESS |
| REVT | SUCCESS | SUCCESS | SUCCESS | SUCCESS |

sequence from randomness. A statistical test is formulated to test a null hypothesis which states that the sequence being tested is random. There is also an alternative hypothesis which states that the sequence is not random. For each test, there is an associated reference distribution (typically normal distribution or $\chi^2$ distribution), based on which a P-value is computed from the bit sequence. If the P-value is greater than a predefined threshold $\alpha$ which is also called significance level, then the sequence would be considered to be random with a confidence of $1-\alpha$, and the sequence passes the test successfully. Otherwise, the sequence fails this test. A P-value of zero indicates that the sequence appears to be completely non-random, and the larger the P-value is, the closer a sequence to a perfect random sequence. In our experiment, we set $\alpha$ to its default value 0.01, which means a sequence passed the test is considered as random with 99% confidence. Before presenting the test results of our proposed three approaches, we would first introduce all 15 statistical tests briefly as follows. A more detailed description for those tests could be found in [2].

The frequency test (FT), the runs test (RT) and the cumulative sum test (CST) are recommended that each sequence to be tested consist of a minimum of $\mathbf{10^2}$ bits (i.e., $\mathbf{n \geq 10^2}$). The frequency Test within a Block (FTB) is recommended that each sequence to be tested consist of a minimum of $\mathbf{M \times N}$ bits (i.e., $\mathbf{n \geq MN}$). The block size M should be selected such that $\mathbf{M \geq 20}$ and $\mathbf{N < 10^2}$. The discrete fourier transform test (DFTT) is recommended that each sequence to be tested consist of a minimum of $\mathbf{10^3}$ bits (i.e., $\mathbf{n \geq 10^3}$). The approximate entropy test (AET) is recommended that each sequence to be tested consist of a minimum of $\mathbf{2^{12}}$ bits (i.e., $\mathbf{n \geq 2^{12}}$). The test for the longest run of ones in a block (LROBT) is recommended that each sequence to be tested consist of a minimum of 6272 bits for M=128. The binary matrix rank test (BMRT) is recommended that each sequence to be tested consist of a minimum of $\mathbf{10^5}$ bits (i.e., $\mathbf{n \geq 10^5}$). The non-overlapping template matching test (NTMT), the overlapping template matching test (OTMT), the maurer's universal statistical test (MUST), the linear complexity test (LCT), the serial test (ST), the random excursions test (RET) and the random excursions variant test (REVT) are recommended that each sequence to be tested consist of a minimum of $\mathbf{2^{20}}$ bits (i.e., $\mathbf{n \geq 2^{20}}$).



(a) Plain image (b) Ciphered image of example of A

Fig. 3. Images of test results.

The NIST suite tests were performed on four bit sequences, each containing $2^{20}$ bits. The P-value as well as final results obtained from the NIST suite for four different sets are given in Table 2. The proposed PRNG successfully passes all randomness tests of NIST suite.

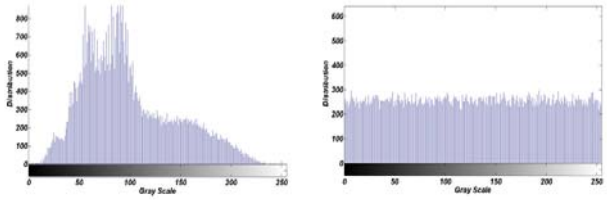
## 4. The proposed encryption scheme and security analysis

### 4.1 Encryption scheme based on pseudorandom bit padding

In the proposed scheme, we create a method to encrypt the image using bits padding. To consider a gray scale image with the size of $\mathbf{M \times N}$ . Here, the plain image is the image of the example of A that an image with the size of $\mathbf{256 \times 256}$ (see Fig. 3(a)). The steps of the encryption are shown below:

- **Step 1:** Generate $\mathbf{8 \times M \times N}$ pseudo-random number sequence using the logistic pseudorandom image algorithm.
- **Step 2:** Transform the image into $\mathbf{8 \times M \times N}$ bit sequence (image sequence).
- **Step 3:** Perform the XOR operation between the image sequence and the pseudo-random bit sequence to form the cipher sequence.
- **Step 4:** Transform the cipher sequence into image matrix I (ciphered image).
- **Step 5:** Divide the matrix I into four parts, uniformly. Move the odd rows with the even rows between the two parts in the main diagonal and between the other two parts, respectively.
- **Step 6:** Divide the matrix I into four parts, uniformly. Move the odd columns with the even columns between the two parts in the main diagonal and between the other two parts, respectively.

The ciphered image is shown in fig. 3(b). The grey scale histograms are given in figs. 4(a), 4(b). The fig. 4(b) shows uniformity in distribution of grey scale of the ciphered images. In addition, the average pixel intensity



(a) Histogram of the plain image (b) Histogram of the ciphered image of example of A

Fig. 4. Histograms of images.

for plain image is 98.92 and for ciphered image is 127.09.

### 4.2 Analysis of security of the proposed encryption scheme

The Security is a major intransitive of a cryptosystem. Here, a complete analysis is made on the security of the cryptosystem. We have tried to explain that this cipher image is sufficiently secure against various cryptographical attacks, as shown below:

#### 4.2.1 Key space analysis

Key space size is the total number of different keys that can be used in the encryption [20]. Security issue is the size of the key space. If it is not large enough, the attackers may guess the image with brute-force attack. If the precision is $\mathbf{10^{-14}}$, the size of key space for initial conditions and control parameters is $\mathbf{2^{306}}$. In addition, we use the black white dynamic images derived of Albert Einstein image with $\mathbf{256 \times 256}$ pixels. The size of the key space for black white dynamic image is no less than $\mathbf{2^{256}}$. This size is large enough to defeat brute-force by any super computer today.

#### 4.2.2 Correlation Coefficient Analysis

The statistical analysis has been performed on the encrypted image from example of A. This is shown by a test of the correlation between two adjacent pixels in plain image and encrypted image. We randomly select 2000 pairs of two-adjacent pixels (in vertical, horizontal, and diagonal direction) from plain images and encrypted images, and calculate the correlation coefficients [19,20], respectively by using the following two equations (see Table 3 and Fig. 5(a) and 5(b)):

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))(y_i - E(y)),$$

$$r_{xy} = \frac{Cov(x,y)}{(D(x))^{\frac{1}{2}}(D(y))^{\frac{1}{2}}}$$

Where

$$E(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i), \quad D(y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2 .$$

Where, E(x) is the estimation of mathematical expectations of x, D(x) is the estimation of variance of x, and Cov(x,y) is the estimation of covariance between x and y, where x and y are grey scale values of two adjacent pixels in the image.

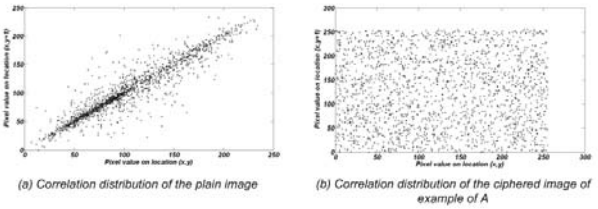



Fig. 5. Correlation distributions of two horizontally adjacent pixels in the plain image and the ciphered image.

Table III. Correlation coefficients of two adjacent pixels in the plain image and the ciphered image of example of A.

| Direction | Plain image | ciphered image |
|---|---|---|
| Horizontal | 0.9341 | 0.0023 |
| Vertical | 0.9634 | 0.0098 |
| Diagonal | 0.9402 | 0.0043 |

### 4.2.3 Differential attack

Attackers try to find out a relationship between the plain image and the cipher image, by studying how differences in an input can affect the resultant difference at the output in an attempt to derive the key [31]. Trying to make a slight change such as modifying one pixel of the plain image, attacker observes the change of the cipher image [31]. To test the influence of one pixel change on the whole encrypted image by the proposed scheme, two common measures are used:

Number of Pixels Change Rate (NPCR) stands for the number of pixels change rate while, one pixel of plain image is changed. Unified Average Changing Intensity (UACI) measures the average intensity of differences between the plain image and ciphered image. The NPCR and The UACI, are used to test the influence of one pixel change on the whole image encrypted by the proposed scheme and can be defined as following:

$$NPCR = \frac{\sum_{i,j} D(i,j)}{W \times H} \times 100\%$$

$$UACI = \frac{1}{W \times H}\left[\sum_{i,j}\frac{C_1(i,j) - C_2(i,j)}{255}\right] \times 100\%$$

where W and H are the width and height of $C_1$ or $C_2$. $C_1$ and $C_2$ are two ciphered images, whose corresponding original images have only one pixel difference and also have the same size. The $C_1(i,j)$ and $C_2(i,j)$ are grey-scale values of the pixels at grid (i,j). The D(i,j) determined by $C_1(i,j)$ and $C_2(i,j)$. If $C_1(i,j) = C_2(i,j)$, then, D(i, j) = 1; otherwise, D(i, j) = 0. We have done some tests on the proposed scheme (256 grey scale image of size $256 \times 256$) to find out the extent of change produced by one pixel change in the plain image. We have obtained NPCR = 0.43% and UACI = 0.34%. The results demonstrate that the proposed scheme can survive differential attack.

## 5. Conclusion

We have proposed a chaotic encryption scheme based on pseudorandom bit padding that the bits be generated by a novel logistic pseudorandom image algorithm. The security of the cipher image of this scheme is evaluated by the key space analysis, the correlation of two adjacent pixels and differential attack. The distribution of the ciphered images is very close to the uniform distribution, which can well protect the information of the image to withstand the statistical attack.

## Appendix

**Monobit Test:**

The goal of this test is to determine whether the frequency of 0's and 1's in bit sequences generated by the PRNG are approximately same [27]. Let $n_0, n_1$ denote the number of 0's and 1's in bit sequences respectively. We calculate $\chi^2$ by using the formula [27]:

$$\chi^2 = \frac{(n_0 - n_1)^2}{n},$$

which approximately follow a $\chi^2$ distribution with one degree of freedom. The computed results are shown in Table 1. The calculated values of $\chi^2$ are less in compared to the critical value of $\chi^2$ at $\alpha = 0.05$ (5% level of significance) and 1df (one degree of freedom). It means that these bit sequences pass the monobit test and can be said to be satisfactorily random with respect to this test [27].

**Serial Test:**

The goal of this test is to determine whether the number of occurrence of pairs 00, 01, 10 and 11 in the bit streams generated by PRNG is approximately same [27]. Let $n_{00}, n_{01}, n_{10}$ and $n_{11}$ denote the number of occurrence of pairs 00, 01, 10 and 11 respectively in the bit sequences. We calculate $\chi^2$ by using the formula [27]:

$$\chi^2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1$$

and the computed values are found to follow approximately the $\chi^2$ distribution with 2 degrees of freedom. The results are shown in Table 2. The calculated values of $\chi^2$ are less than critical value of $\chi^2$ at $\alpha = 0.05$ (5% level of significance) and 2df (two degrees of freedom). It means that bit sequences pass the serial test and are satisfactorily random with respect to this test.

## References

[1] N. K. Pareek, V. Patidar, K. K. Sud, Image encryption using chaotic logistic map, Image and Vision Computing 24 (2006) 926–934.
[2] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel,D. Banks, A. Heckert, J. Dray, S. VoA, statistical test suite for random and pseudorandom number generators for cryptographic applications, NIST special publication 800-22 (2010) 5-125.
[3] G. Jakimoski, L. Kocarev, Block encryption ciphers based on chaotic maps, IEEE Transaction on Circuits System-I. 48 (2002) 163-169.
[4] R.L. Rivest, A. Shamir, L.M. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Communications of the ACM. 21 (1978) 120-126.
[5] M. Emre-Celebi, Distance measures for reduced ordering-based vector filters, IET Image Processing. 3 (2009) 249-260.
[6] B. Wang, Q. Wu, Y. Hu, A knapsack-based probabilistic encryption scheme, Information Sciences. 177 (2007) 3981-3994.
[7] F. Cao, Z. Cao, A secure identity-based proxy multi-signature scheme, Information Sciences. 3 (2009) 292-302.
[8] D. Xiao, X. Liao, S. Deng, A novel key agreement protocol based on chaotic maps, Information Sciences. 177 (2007) 1136-1142.
[9] D. Xiao, X. Liao, S. Deng, Using time-stamp to improve the security of a chaotic maps-based key agreement protocol, Information Sciences. 178 (2008) 1598-1602.
[10] R. M. DSouza, Y. Bar-Yam, M. Kardar, Sensitivity of ballistic deposition to pseudorandom number generators, Phys. Rev. E. 57 (1998) 5044-5052.
[11] J. F. Fernandez, C. Criado,Algorithm for normal random numbers, Phys. Rev. E. 60 (1999) 3361-3365.
[12] I. Vattulainen, T. Ala-Nissila, K. Kankaala,Physical models as tests of randomness, Phys. Rev. E. 52 (1995) 3205-3214.
[13] Q. Zhou, X. Liao, K.W. Wong, Y. Hu, D. Xiao, True random number generator based on mouse movement and chaotic hash function, Information Sciences. 179 (2009) 3442-3450.
[14] H. Lu, D. Wang, R. Zhang, Y.-W. Chen, Video object pursuit by tri-tracker with on-line learning from positive and negative candidates, IET Image Processing. 5 (2011) 101-111.
[15] J. Fridrich, Symmetric ciphers based on two-dimensional chaotic maps, International Journal of Bifurcation and Chaos. 8 (1998) 1259-1264.
[16] Q. Zhou, K.W. Wong, X. Liao, T. Xiang, Y. Hu, Parallel image encryption algorithm based on discretized chaotic map,Chaos Solitons Fractals. 38 (2008) 1081-1092.
[17] R. Tenny, L.S. Tsimring, Additive mixing modulation for public key encryption based on distributed dynamics, IEEE Transactions on Circuits and Systems-I. 52 (2005) 672-679.
[18] Y. Wang, X. Liao, K. Wong, One-way hash function construction based on 2D coupled map lattices, Information Sciences. 178 (2008) 1391-1406.
[19] G. Chen, Y. Mao, C.K. Chui, A symmetric image encryption scheme based on 3D chaotic cat maps, Chaos Solitons Fractals 21 (2004) 749-761.
[20] S. Behnia, A. Akhshani, S. Ahadpour, H. Mahmodi, A. Akhavan, A fast chaotic encryptions cheme based on piece wise nonlinear chaotic maps, Physics Letters A 366 (2007) 391–396.
[21] X. Yi, Hash function based on chaotic tent maps, IEEE Transactions on Circuits and Systems-II. 52 (2005) 354-357.
[22] J. Zhang, X. Wang, W. Zhang, Chaotic keyed hash function based on feedforward-feedback nonlinear digital filter, Physics Letters A. 362 (2007) 439-448.
[23] S. C. Phatak S. Suresh Rao, Logistic map: A possible random-number generator, Phys. Rev. E. 51 (1995) 3670-3678.
[24] C. Peng, S. Prakash, H. J. Herrmann, H. E. Stanley, Randomness versus deterministic chaos: Effect on invasion percolation clusters, Phys. Rev. A. 42 (1990) 4537-4542.
[25] A. Kanso, N. Smaoui, Logistic chaotic maps for binary numbers generations, Chaos, Solitons and Fractals. 40 (2009) 2557-2568.
[26] N. Smaoui , E. Kostelich Using chaos to shadow the quadratic map for all time, Int J Comput Math. 70 (1998) 117-129.
[27] N. K Pareek, V. Patidar, K. K Sud, A Random Bit Generator Using Chaotic Maps, International Journal of Network Security. 10 (2010) 32-38.
[28] H. Javashi and R. Sabbaghi-Nadooshan, A Novel Elliptic curve cryptography Processor using NoC design, IJCSI International Journal of Computer Science Issues. 8, 3 (2011) 376-381.

[29] P. Balakumar and R. Venkatesan, Secure Biometric Key Generation Scheme for Cryptography using Combined Biometric Features of Fingerprint and Iris , IJCSI International Journal of Computer Science Issues. 8, 5 (2011) 349-356.
[30] **S. Kandar, A. Maiti, B. C. Dhara,** Visual Cryptography Scheme for Color Image Using Random Number with Enveloping by Digital Watermarking, IJCSI International Journal of Computer Science Issues. 8, 3 (2011) 543-549.
[31] Sh. R. Maniyath and M. Supriya, An Uncompressed Image Encryption Algorithm Based on DNA Sequences, Computer Science and Information Technology (CS , IT). 2 (2011) 258–270. (DOI: 10.5121/csit.2011.1224)

**Sodeif Ahadpour** is a assistant professor at the University of Mohaghegh Ardabili, Ardabili, Iran. He has obtained his Master degree in Atomic and Molecular Physics (1993) and his PhD in Theoratical Physics (2007), both in Tabriz university, Tabriz, Iran. His research interest includes quantum computing & quantum information, discrete chaotic dynamical systems, graph theory & complex network, chaotic cryptography, wavelet, random generator.

**Yaser Sadra** is a research scientist at the the University of Mohaghegh Ardabili, Ardabili, Iran. He has obtained his Master degree in Theoratical Physics (2008) from the University of Mohaghegh Ardabili, Ardabili, Iran. His research interest includes graph theory & complex network, complex systems analysis, markov chain, cryptography, random generator.

**Zahra ArastehFard** is a research scientist at the the University of Mohaghegh Ardabili, Ardabili, Iran. He has obtained his Master degree in Organic Chemistry (2010) from the University of Mohaghegh Ardabili, Ardabili, Iran. His research interest includes synthesis, conformational analysis, markov chain, cryptography.

# Control Logic Algorithm for Medium Scale Wind Turbines

**O.H.Abdel Satar[1], Saad.M.A.Eid[2], E.M.Saad[3], R.R.Darwish[4]**

**[1]Faculty of Engineering, Helwan, Cairo, Egypt**

**[2]Faculty of Engineering, Cairo, Cairo, Egypt**

**[3]Faculty of Engineering, Helwan, Cairo, Egypt**

**[4]Faculty of Engineering, Helwan, Cairo, Egypt**

**Abstract**

Recently, sustainable attention has been drawn to renewable energy sources. Wind energy systems as renewable source of energy have been extensively studied because of its benefits as an environmentally friendly clean energy, inexhaustible, safe and a low-cost for long term. Because of its unpredictable availability, power management control algorithms are essential to extract as much power as possible from the wind during its availability durations. This paper is motivated for proposing the main control algorithm for wind turbines each incorporating two generators. The proposed main algorithm contains several sub algorithm models (strategies) for power control, pitch control, status checking, starting, grid connection, normal and emergency shutdown that are studied, designed and also, tested under operation. The testing phase shows that in the high wind speed range, the pitch control seems the most relevant to release a power margin. While in the low wind speed range, the increase of the rotation speed is more convenient.

***Keywords:*** *wind turbine, control logic algorithm, pitch control, status checking, starting, grid connection, normal shut down, and emergency shutdown*

## 1. Introduction

Wind energy has been harnessed by many generations for thousands of years to mill grain, pump water and sailing [5]. Just in last decade, the wind energy industry has experienced a growth of almost 30 percent each year [1]. Due to the increasing concern about the environment and the depletion of natural resources such as fossil fuels, much research is now focused on obtaining new environmentally friendly sources of power. The lack of accurate models must be countered by robust control strategies capable of securing stability and some performance features despite model uncertainties. The controller measures the following parameters as analogue signals Voltage on all three phases, Current on all three phases, Frequency on one phase, Temperature inside the nacelle, Generator temperature, Gear oil temperature, Gear bearing temperature, Wind speed, The direction of yawing, Low-speed shaft rotational speed and High-speed shaft rotational speed. The controller also measures the following parameters as digital signals Wind direction, Over-heating of the generator, Hydraulic pressure level, Correct valve function, Vibration level, Twisting of the power cable, Emergency brake circuit, Brake-caliper adjustment, Overheating of small electric motors for the yawing, hydraulic pumps, etc. by several sensors[21]. The control problems are even more challenging when turbines are able to operate at variable speed and variable pitch. The best use of this type of turbine can only be achieved by means of multivariable controllers. However, due to wind's erratic nature, intelligent control strategies must be implemented to harvest as much potential wind energy as possible while it is available. Because of its advantages, erratic nature, and recent technological advancements in wind turbine aerodynamics and power electronic interfaces, wind energy is considered to be an excellent supplementary energy source. Research to extract the maximum power out of wind energy is an essential part of making wind energy much more viable and attractive. In addition to increasing the energy capture, wind turbines can be controlled to reduce the loading on the drive-train and tower structure, leading to potentially longer installation life. Increasingly, modern wind turbines include mechanical actuators with the aim of having control of the blade pitch angle [16]. Pitch control is commonly meant to limit the captured power above rated wind speed, bringing about more cost-effective designs.

## 2. Background

The equations of kinetic energy is given by Eq. (1) and power describe the potential energy that can be harnessed from the wind is described by Eq. (2) [6].

$$EK = 0.5mv_m = 0.5\rho AdV^2_w \quad (1)$$

$$P_w = 0.5mv_m / t = 0.5\rho AdV^2_w / t = 0.5\rho AV^3_w \quad (2)$$

A turbine's efficiency, and thus power coefficient ($C_p$) curve, is what differentiates one turbine from another. By taking the efficiency of the turbine into account, Eq. (3) represents the mechanical power captured by the wind by any turbine. The power coefficient can be evaluated by Eq. (4). From equation (3), it can be observed that the

$$P_w = 0.5\rho A\, C_p(\beta, \lambda)\, V^3_w \quad (3)$$

power available in the wind is proportional to the cube of the wind speed. This means that there is much more energy in high speed winds than in slow winds.

$$C_p(\beta, \lambda) = \text{actual turbine power} / \text{theoretical turbine power} = P_m / P_w = P_m / 0.5\rho AV^3_w \quad (4)$$

The mathematical representation of the tip speed ratio is given to be as follows in Eq. (5) [6].

$$\lambda = R\,\omega_b / V_w \quad (5)$$

Where ρ = air density, A = rotor swept area, d = distance, m = mass of air =air density * volume = ρ*A*d, $V_w$ = distance/time, $C_p(\beta, \lambda)$ = power coefficient function, λ = the tip speed ratio, and β = pitch angle.

To maximize the amount of power captured by the turbine, variable-speed wind turbine systems are used because they allow turbine speed variation ([2],[3],[4],[7],[8],[9],[10],[11]).

### 2.1 Control Strategy

The primary challenge of wind energy systems is to be able to capture as much energy as possible from the wind in the shortest time. From the electronics point of view, this goal can be achieved through different converter topologies and maximum power. The main control goals were the limitation of power and speed below some specified values to prevent the turbine from unsafe operation under high wind conditions. The control systems have been expected not merely to keep the turbine within its safe operating region but also to improve efficiency and quality of power conversion. The development of a wind turbine control system can be divided into several tasks. The first task is to define clearly the control objectives. The second task is the selection of a suitable control strategy, which settles the operating point of the turbine for each wind speed. The third task is to decide how the control strategy will be realized. It encompasses the selection of the control schemes, the controlled variables, the reference signals, the switching procedure between different controllers, etc. This step is usually referred to as controller setup. Finally the last task previous to the implementation is the design of the input-output map, i.e., the dynamic characteristics of the controller according to the specifications.

### 2.2 Controller Objectives

A wind turbine is essentially a device that captures part of the wind energy and converts it into useful work. In particular, Wind Energy Conversion System (WECS) connected to electric power networks must be designed to minimise the cost of supplied energy ensuring safe operation as well as acoustic emission and power quality standards [12]. The minimisation of the energy cost involves a series of partial objectives (Energy capture, Mechanical loads, Power quality).These primary objectives (partial goals) of wind turbine control systems can be arranged in the following topics.

#### 2.2.1 Energy capture

Maximisation of energy capture taking account of safe operation restrictions such as rated power, rated speed and cut-out wind speed, etc. For a wind turbine, the generation capacity specifies how much power can be extracted from the wind taking into consideration both physical and economic constraints. It is usually represented as a curve on the generated power – wind speed plane, the so-called ideal power curve. The ideal power curve for a typical wind turbine is sketched in Figure 1. The range of operational wind speeds is delimited by the cut-in ($V_{min}$) and cut-out ($V_{max}$) wind speeds. The turbine remains stopped beyond these limits. Below cut-in wind speed, the available wind energy is too low to compensate for the operation costs and losses. Above cut-out wind speed, the turbine is shut down to prevent from structural overload. Even though wind speeds above $V_{max}$ contain huge energy, their contribution to the annual average energy is negligible.

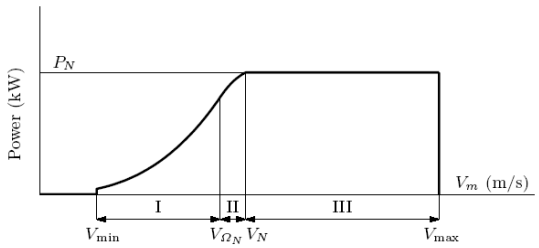



Fig. 1: Ideal power curve

This is corroborated by Figure 2 where a typical power density function at a given site is outlined. It is observed there that the energy left to be captured because of keeping the turbine stopped beyond the wind speed limits $V_{min}$ and $V_{max}$ is comparatively low. It can also be noted in Figure 1 that the ideal power curve remains constant at rated power

$P_N$ above wind speed $V_N$ named rated wind speed $V_N$. For instance, designing the turbine to extract all the available energy up to cut-out wind speed would lead to an increment in the cost per kW. The ideal power curve exhibits three different regions with distinctive generation

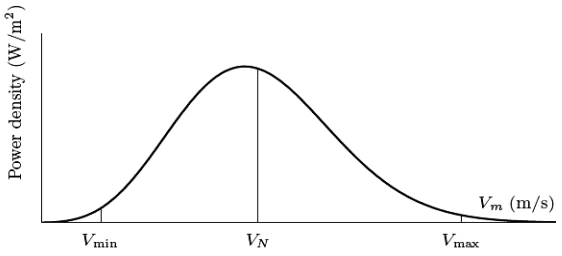



Fig. 2: Power density vs. wind speed

objectives. At low wind speeds (region I), the available power is lower than rated power. The available power is defined as the power in the wind passing through the rotor area multiplied by the maximum power coefficient $C_{Pmax}$, that is So, the generation objective in region I is to extract all the available power. Therefore, the ideal power curve in this region follows a cubic parabola defined by Eq. (6). On

$$P_{av} = C_{Pmax}P_w = 0.5\rho\pi R^2 C_{Pmax} V^3 \quad (6)$$

the other side, the generation goal in the high wind speed region (region III) is to limit the generated power below its rated value to avoid over loading. In this region the available power exceeds rated power; therefore the turbine must be operated with efficiency lower than $C_{Pmax}$. Finally, there is region II, which is actually a transition between the optimum power curve of region I and the constant power line of region III. In this region, rotor speed is limited to maintain acoustic noise emission within admissible levels and to keep centrifugal forces below values tolerated by the rotor. Eventually, in the case that such a speed limit is not reached, region II may not exist and the optimum power curve (i.e., region I) may continue until getting to rated power. So the control strategy settles the steady-state values of torque (or power) and rotor speed for each wind speed within the range of turbine operation. The control strategy affects the controller setup and design. In fact, the control schemes may differ from one region of operation to another.

### 2.3.2 Mechanical loads

Mechanical loads preventing the WECS from excessive dynamic mechanical loads. This general goal encompasses transient loads alleviation, high frequency loads mitigation and resonance avoidance. Considering the minimisation of the energy cost, the control system should not merely design to track as tightly as possible the ideal power curve. The mechanical loads wind turbines are exposed to must also be considered [13] because it may cause fatigue damage on several devices, thereby reducing the useful life of the system. The transition between maximum power tracking (region I) and power regulation (region III) and the way power is limited in above rated wind speeds have a direct impact on transient loads. Unsuitable control strategies may inevitably lead to strong transient loads. Cyclic loads are highly influenced by the control strategy as well as by the controller setup and design. The control of the electric generator affects the propagation of drive-train loads whereas the pitch control impacts directly on the structural loads. Therefore, inappropriate control designs might accentuate the vibration modes, potentially leading to the destruction of some mechanical devices such as gearbox or blades. The controller must provide damping at the vibration modes whenever possible in order to mitigate high frequency loads and reduce the risk of fatigue breakdown. So, the control strategy must avoid operation at points where those vibration modes that cannot be damped by the controller are likely to be excited ([14], [12]).

### 2.2.3 Power quality

Power quality conditioning the generated power to comply with interconnection standards to smooth the power supplied to the grid. It affects the cost of energy in several ways. Poor power quality may demand additional investments in power lines, or may impose limits to the power supplied to the grid. The control system design must also take power conditioning into account. This control requirement is more and more relevant as the power scale of wind generation facilities approaches the output rating of conventional power plants [34]. Power quality is mainly assessed by the stability of frequency and voltage at the point of connection to the grid and by the emission of flicker [15].The interaction of wind turbines with the power network affects the voltages at the grid terminals. On the one hand, slow voltage excursions take place when the power extracted by the WECS changes with mean wind speed. The amplitude of these variations closely depends on the impedance of the grid at the connecting point and on the active and reactive power flow. Reactive power, power factor or, directly, voltage regulation can be accomplished by an adequate control of the electronic converters ([16],[ 17]). The voltage fluctuations and flicker can be attenuated by including passive or active filters, or by controlling the reactive power handled by the electronic converters ([18],[19],[20]). Also, they can be smoothed indirectly by tackling the propagation of the cyclic loads. This is achieved incorporating dynamic damping to the drive-train by means of a suitable control of the generator torque characteristic [12]. These objectives are actually closely related and sometimes conflicting. Therefore, they should not be pursued separately. Conversely, the control target is to find a well balanced compromise among them. Some control strategies are designed to maximize the energy

extraction; others accept a reduced energy capture in order to avoid operating regions where heavy mechanical loads are being inevitable.

## 3. The Proposed Control Logic Algorithm

The proposed algorithms, depicted in figure (3), comprise several modules .Those modules are hydraulic pump, tip brake, disc brake, generator etc. All these modules are controlled by main control. Main control manages when to apply the disc brake, pull in the tip brake or connect the generator to the grid.

### 3.1 Main control module

The main control module is divided into four main states as described in figure (3). Each state indicates generally what the turbine is doing. Each main state contains several sub states. The sub states describe more specifically what the turbine is doing. It could either produce on generator 1(G1), change from (G1) to generator 2(G2) or using emergency Brake programme to stop the turbine.

### 3.2 Main states

Figure (3) shows the four main states. The arrow shows how the proposed programme goes from one main state to another. One thing special is that from Start to Operate, the proposed programme can go directly to Braking. If a stop condition becomes active it can interrupt any running sub state, No matter which sub state is running. The stop main state (Stopped)

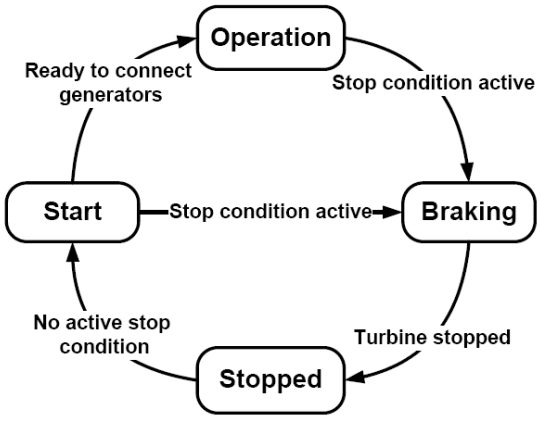



Fig. 3: The main control module

contains all the sub stop states which handle all stop conditions as Emergency stop, grid failure etc. Start of wind turbine (Start state). Self tests and start sequence are performed. Brakes down the wind turbine (Braking state) by applying the mechanical disc brake or releases the tip brakes. The rotor speed will decrease and when the rotor reaches zero rpm, the programme will continue to main state Stopped. The wind turbine is producing power on either the G2 or G1 (Operate state).

## 4. Control logic general description

Figure (4) describes the developed state diagram for the control logic of the proposed wind turbine controller. The Protection functions related to grid errors, temperatures and over speed are implemented to ensure protection of the turbine mechanical and electrical system. The normal operation of the wind turbine will be discussed including normal start up and shut down. Then the emergency situations will be described and finally wind turbine reactions to previously stated models will be described. The off operation can be generally classified into temporary and stationary. In Temporary modes the wind turbine is meant to be on them for a limited period of time. These modes are status checking, starting, grid connection, normal shut down and emergency shutdown. In The Stationary modes the wind turbine can be on for an unlimited period of time. The Stationary modes include stop and normal operation (partial load and full load). In the following, subsections the state diagram mode are described.

### 4.1 Stop mode

In this mode the generator is disconnected from the grid, the blades are 90 degrees to the rotor plane (pitch control) and mechanical brake on. This status can be achieve as a normal stop for example when there is no wind, after manual stop (by the operator), or after an emergency stop. In the last two cases the wind turbine will not change its mode until is reset by operator.

### 4.2 Status checking

After the turbine has been order to start, the controllers checks the status of all subsystem and reads all measurable variables, and then check that this values are within the acceptable range. During this process if any error sign appears, the preposed controller invalidates any other mode.

### 4.3 Starting

After the status has been checked and no error sign appears and the average wind speed match the desired one, the mechanical brake is released and the blades are pitched the rotor will start turning and when the connection conditions are achieve (normally it means a certain rotor speed), the grid connection process start

### 4.4 Grid connection

The grid connection has normally two steps soft connection and normal connection. The soft connection is used for few seconds just to avoid the high current and high mechanical loads. Then after the equilibrium condition is reached the connection shift to the normal connection.

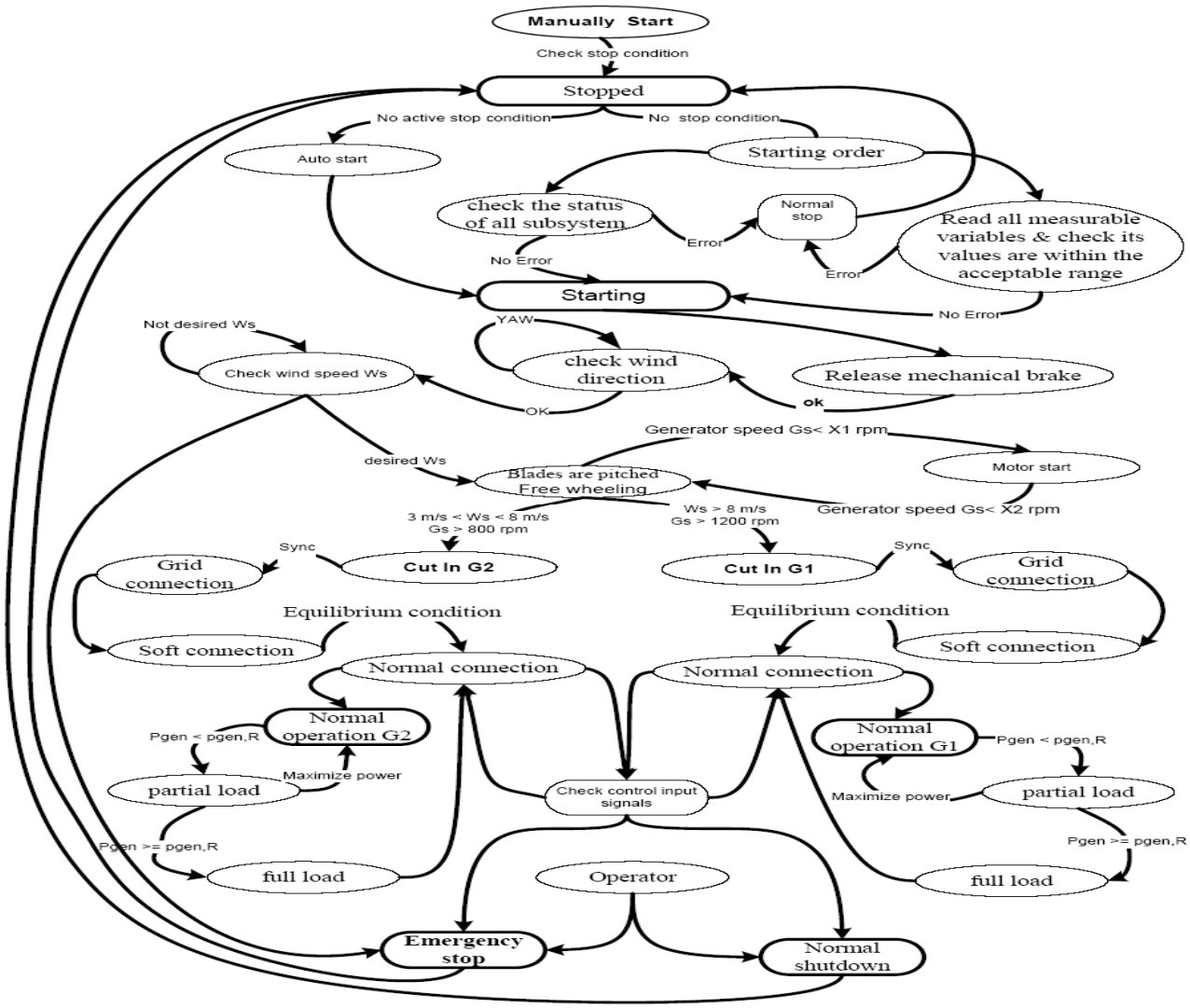



Fig. 4: The state diagram control logic

### 4.5 Normal operation

As discussed previously the turbine operation include two main power areas (partial load and full load).each power area has a distinct objective to be fulfilled by the control. In the pre rated power area the main objective is to maximize power production. While the post rated power area the main objective turn as into keeping the power under control as well as the loads on the machinery and the electrical part.

### 4.6 Normal shutdown

Depending on the design there could be several reason for a normal shutdown one of the most common are the low wind condition, by operator, and some control variable that is out of limit like oil temperature, excessive wind, etc. If one of the above mentioned conditions in detected by the controller the blades pitch start normal stop mode.

### 4.7 Emergency stop

Emergency stop occurs in the same way as the normal shutdown but faster. Some examples of event that can result in this kind of stop are big vibrations, over speed of the generator.

## 5. Main Control algorithm

The required algorithm can then be formulated according to the developed state diagram into the following main algorithm, Figure (4). This subsection considers an illustration to describe the operation of the proposed controller. The main control algorithm for the wind turbine is sketched in Figure (5).The control proposed algorithm designed through understanding how all operation state of

wind turbine system is combined. an imaginary situation is consider where the wind will rise from 0 till 25 m/s in steady steps longer that 10 min. and try to follow the control procedures by using the ideal power curve sketched in Figure (1)..While the wind speed is below $V_{min}$ (3) m/s the system will not issue any order. When $V_{min}$ (3) m/s is reached then the status checking is performed. If there is an error sign the stop order will be issue. If all parameter are ok the yaw system is activated and the wind turbine start to keep tracking of the wind direction but the brake is still on because there is not enough wind to produce power effectively. When the wind gets to 4 m/s the mechanical brake is release and the blades are Pitch to the right angle (pitch controlled) so the rotor start turning by itself and accelerates until it reaches the synchronous condition then the grid connection mode takes over and the soft connection is activated first and after few seconds the main contactor for the grid connection is activated. In cases

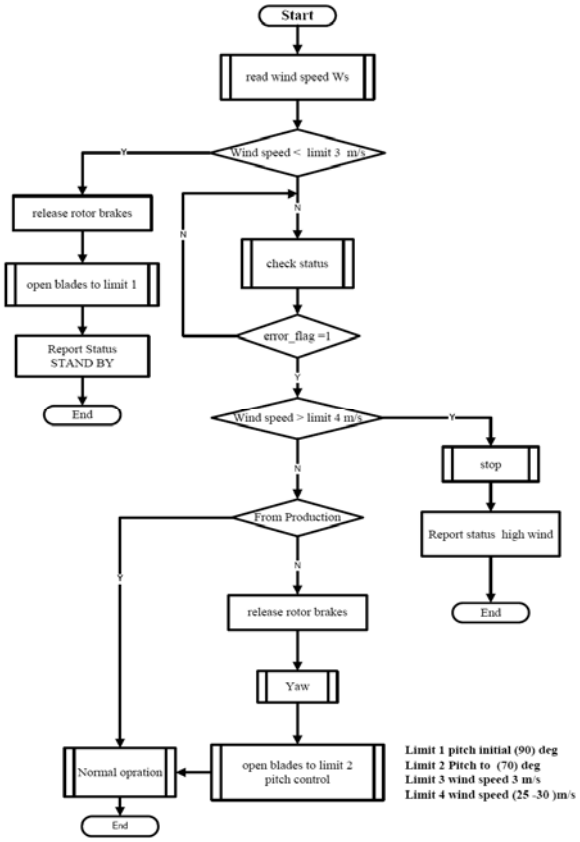



Fig. 5 The main control algorithm for wind turbine

of a wind turbine with a two generators, the small generator will be trigger first. This generator is slower than the bigger one (2/3) and the power is also smaller (8/27).In other cases will result in a set of control parameters for the pitch angle and maximum torque. The turbine is now connected to the grid in the normal operation mode Figure (6) and in the partial load area. In the case of variable pitch and variable speed turbine adjustment of parameter (pitch angle, rotor speed) will be needed to maximize power production. When the power reaches to a value over 80 % of the nominal power of the small generator the conditions are given to trigger the bigger generator. This process cannot

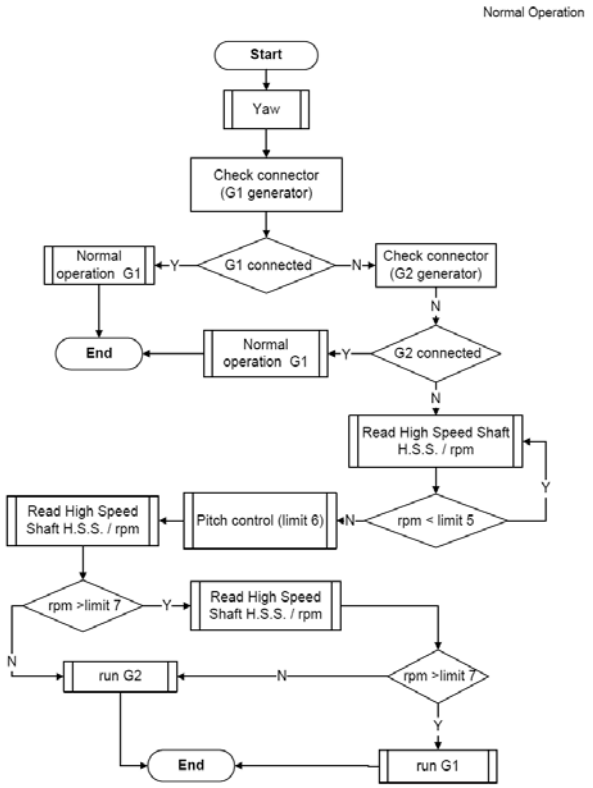



Fig. 6 The normal operation algorithm

be done directly because the speed of the two generator are different and the second ones in higher than the first one so if we connected directly it will implicated a big consumption of energy and big loads in the mechanical parts and the blades. So the normal procedure will be to set the rotor free and allowed to accelerate by itself and when we reach the new synchronous conditions the grid connection process starts again. Any ways the control box will keep the surveillance of the operation and safety parameter. This last mode will continue as long as the wind speed is below $V_{max}$ 25 m/s. when the wind speed goes over 25 m/s the turbine receive the order to change to normal shutdown modes. Then blades are pitch at (3-6) degree per second toward the wind and the generator is disconnected from the grid after the rotor have reduced the speed to zero the mechanical brake is applied.

## 6. Results

The control Algorithm was tested on a wind turbine through separate stages for the following functions: automatically directing the wind turbine in wind direction (yawing), Removing the era of automatic cable (cable twist), the normal stop and emergency shutdown of the

turbine, and other jobs done manually, such as, Guide the turbine to the right and left, Control the opening angle of blade (from angle 2 ° to 90 ° angle), Stop the turbine and run manually, response of the turbine failures. All the previous functions successfully tested individually and then collected together, Nacelle is automatically directed to face the wind according to wind direction sensor signal, and also manually through the unit of the (hand held). The wind turbine is automatically stopped at the arrival of signal cables to the era then the controller directs the Nacelle automatically (YAW) in the opposite direction to lift the age of the cables. Finally the normal operation algorithm tested successfully and the turbine produced power and changing from generator 1(G1) to generator 2(G2) according to wind speed.

## 7. Conclusion

Due to wind's unpredictable nature, power management concepts are necessary to extract as much power as possible from the wind when it becomes available. The proposed algorithm has been developed to maintain the system at its highest possible efficiency by using its memory feature to infer the optimum parameters for wind turbine that have not occurred before. Another feature of the proposed algorithm is that it can be easily customized for various wind turbines since it is independent of turbine characteristics. The proposed algorithm uses a modified version of an algorithm that tested on 20 / 100 kW wind turbine in Hurghada area . This algorithm is characterized as the most appropriate for isolated electrical network (small and medium-sized wind power with or without battery) or for the electrical network connected to the network. The proposed algorithm has gleaned insight into the practical considerations of design control systems for wind turbines. A distinction has been made between supervisory and safety control and separate control issues identified. Common control systems and methods for implementation of these systems in modern wind turbines have been examined. The approach to controlling a wind turbine may vary but the primary objectives remain the same.

## REFERENCES

[1] "Global wind 2006 report," June 2007.

[2] M. Idan, D. Lior, and G. Shaviv, *A robust controller for a novel variable speed wind turbine transmission*, Journal of Solar Energy Engineering, vol. 120,pp. 247–252, 1998.

[3] J. Marques, H. Pinheiro, H. Grundling, J. Pinheiro, and H. Hey, *A survey on variable-speed wind turbine system,* Proceedings of Brazilian conference of electronics of power, vol. 1, pp. 732 – 738, 2003.

[4] Q. Wang and L.-C. Chang, *An intelligent maximum power extraction algorithm for inverter-based variable speed wind turbine systems*, IEEE Transactions on Power Electronics, vol. 19, pp. 1242–1249, Sep.2004.

[5] T. I. of Electrical and E. E. Inc, IEEE Canadian review: Green power, *Mathematical and Computational Applications*, pp. 10–17, December 2007.

[6] F.-S. dos Reis, K. Tan, and S. Islam, *Using pfc for harmonic mitigation in wind turbine energy conversion systems*, IEEE Industrial Electronics Society Conference, pp. 3100–3105, November 2004.

[7] G. Moor and H. Beukes, *Power point trackers for wind turbines*, Power Electronics Specialist Conference (PESC), pp. 2044–2049, 2004.

[8] Y. Song, B. Dhinakaran, and X. Bao, *Variable speed control of wind turbines using nonlinear and adaptive algorithms,* Journal of Wind Engineering and Industrial Aerodynamics.,vol. 85, no. 3, pp.293–308, 2000.

[9] D.-S. Zinger, *Annualized wind energy improvement using variable speeds*, IEEE Transactions on Industry Applications, vol. 33, pp. 1444–1447, November 1997.

[10] M. Chinchilla, *Control of permanent-magnet generators applied to variable-speed wind-energy systems connected to the grid*, IEEE Transactions on Energy Conversion, vol. 21, pp. 130–135, March 2006.

[11] E. Koutroulis and K. Kalaitzakis, *Design of a maximum power tracking system for wind-energy-conversion applications*, IEEE Transactions on Industrial Electronics, vol. 53, April 2006.

[12] Leithead, W. and Connor, B. (2000). *Control of variable speed wind turbines: design task*. International Journal of Control 73(13), 1189– 1212.

[13] De Battista, H., Mantz, R., and Christiansen, C. (2003). *Energy-based approach to the output feedback control of wind energy systems*. International Journal of Control 76(3), 299–308.

[14] Jauch, C., Matevosyan, J., Ackermann, T., and Bolik, S. (2005). *International comparison of requirements for connection of wind turbines to power systems*. Wind Energy 8(3), 295–306.

[15] Muljadi, E., Butterfield, C., Chacon, J., and Romanowitz, H. (2006).*Power quality aspects in a wind power plant*. Technical Report NREL/CP-500-39183, National Renewable Energy Laboratory, Golden, USA.

[16] Hansen, M., Hansen, A., Larsen, T., S., Sǿrensen, P., and Fuglsang, P. (2005). *Control design for a pitch-regulated, variable speed wind turbine*.Technical Report RISO-R-1500(EN), RISO National Laboratory, Roskilde, Denmark.

[17] Tapia, A., Tapia, G., and Ostolaza, J. (2004). *Reactive power control of wind farms for voltage control applications*. Renewable Energy 29, 377–392.

[18] Larsson, A. (2002). *Flicker emission of wind turbines caused by switching operations*. IEEE Transactions on Energy Conversion 17(1), 119–123.

[19] Sun, T., Chen, Z., and Blaabjerg, F. (2005). *Flicker study on variable speed wind turbines with doubly fed induction generators*. IEEE Transactions on Energy Conversion 20(4), 896–905.

[20] Thiringer, T., Petru, T., and Lundberg, S. (2004). *Flicker contribution from wind turbine installations*. IEEE Transactions on Energy Conversion 19(1), 157–163.

[21] Rules and Guidelines Industrial Services , Guideline for certification of wind turbine, Germanischer Lloyd WindEnergie GmbH-2003

**Osama Abdel Hakeem Abdel Sattar** is a Production Research Center Manager in A.O.I.E.F. He received his B.Sc., from Faculty of Engineering, Department of Electrical Engineering, (Electronics and Communication section) Helwan University, Egypt, his Dipl.-Ing. degree (in Control Eng.) and M.Sc.( in computer engineering) in 1993, 2001 and 2005,respectively. He became a Teacher in the Ministry Of Education in 1997, and R&D Engineer in A.O.I in 2000, and VLSI Designer in A.O.I in 2005, and wind turbine Designer in A.O.I in 2008 until now. He is a member of the Society and HF Technology, DigChip Member ,DriverGuide Member.

**R. R. Darwish** received the B.Sc. and M.Sc degrees in Electronics and Communications Engineering from Helwan University, Egypt in 2000 and 2004, respectively. She has obtained her Ph.D in the field of wireless sensor network from the Department of Electronics and Communications Engineering at Helwan University in 2009. Currently she is an associative professor in the Department of Mechatronics at the Faculty of Engineering, Helwan University, Egypt. Her research interests include image processing, wireless sensor networks, and cloud computing.

**Saad Mohamed Ali Eid** is a Professor of Control Engineering, Faculty of Engineering, Univ. of Cairo. He received his B.Sc. degree and M.sc. degree in Electronics Engineering (Communication section) from Cairo Univ., his Dr.-Ing degree from Stuttgart Univ. , West Germany ,at 1963,1968 and 1973 respectively . He became an Associate Prof. and a Professor in 1978, and 1983 respectively. He was an International scientific member of the ECCTD, 1983. He is Author of 70 and/or Co-author of 200 scientific papers.

**Elsayed Mostafa Saad** is a Professor of Electronic Circuits, Faculty of Engineering, Univ. of Helwan. He received his B.Sc. degree in Electrical Engineering (Communication section) from Cairo Univ., his Dipl.-Ing. degree and Dr.-Ing degree from Stuttgart Univ. , West Germany ,at 1967,1977 and 1981 respectively . He became an Associate Prof. and a Professor in 1985, and 1990 respectively. He was an International scientific member of the ECCTD, 1983. He is Author and/or Co-author of 132 scientific papers. He is a member of the national Radio Science Comittee, member of the scientific consultant committee in the Egyptian Eng. Syndicate for Electrical Engineers, till l May 1995, Member of the Egyptian Eng. Sydicate, Member of the European Circuit Society (ECS), Member of the Society of Electrical Engineering(SEE).

# Adaptive and Reliable Control Algorithm for Hybrid System Architecture

**O.H.Abdel Satar[1], Saad.M.A.Eid[2], E.M.Saad[3], R.R.Darwish[4]**

**[1]Faculty of Engineering, Helwan, Cairo, Egypt**

**[2]Faculty of Engineering, Cairo, Cairo, Egypt**

**[3]Faculty of Engineering, Helwan, Cairo, Egypt**

**[4]Faculty of Engineering, Helwan, Cairo, Egypt**

**Abstract**

A stand-alone system is defined as an autonomous system that supplies electricity without being connected to the electric grid. Hybrid systems combined renewable energy source, that are never depleted (such solar (photovoltaic (PV)), wind, hydroelectric, etc.) , With other sources of energy, like Diesel. If these hybrid systems are optimally designed, they can be more cost effective and reliable than single systems. However, the design of hybrid systems is complex because of the uncertain renewable energy supplies, load demands and the non-linear characteristics of some components, so the design problem cannot be solved easily by classical optimisation methods. The use of heuristic techniques, such as the genetic algorithms, can give better results than classical methods. This paper presents to a hybrid system control algorithm and also dispatches strategy design in which wind is the primary energy resource with photovoltaic cells. The dimension of the design (max. load) is 2000 kW and the sources is implemented as flow 1500 kw from wind, 500 kw from solar and diesel 2000 kw. The main task of the preposed algorithm is to take full advantage of the wind energy and solar energy when it is available and to minimize diesel fuel consumption.

***Keywords:*** *wind turbine, economic control algorithm, dispatch strategy, hybrid system, renewable energy sources, photovoltaic cells (PV), genetic algorithms.*

## 1. Introduction

The economic dispatch is a significant function in the modern energy system [1]. It consists in programming correctly the electric production in order to reduce the operational cost ([2], [3]). The economic dispatch problem can be formulated as a multi objective optimization problem ([4],[5],[6],[7],[8]). It includes in hybrid systems to distribute the renewable productions between the Diesel power stations by the most economic way, to reduce the emissions of the polluting gases and to maintain the stability of the network after penetration of renewable energy. This production poses many technical problems for their integration in the electric system. The number of decision variables of the problem is related to all the nodes of the network (diesel power, wind power and solar power). The control system is subject to the specific constraints of a particular application. Hybrid energy systems are recognised as a viable alternative to reticulated grid supply or conventional, fuel-based, remote area power supplies [10]. The design and operation control [9] is not a linear problem due to non-linear component characteristics with a large number of variables [11]. The optimal design of problems like this cannot be achieved easily using classical optimisation methods. This paper presents a method of optimisation economic dispatch for Wind-PV-Diesel systems using a Genetic Algorithm (GA). The proposed Architecture for hybrid system is shown in Figure 1. There are some programs that simulate hybrid systems, as HYBRID2 [12], and TRNSYS [19]. HYBRID2 simulates hybrid systems with very high precision calculations, but it does not optimise the system. TRNSYS was initially developed to simulate thermal systems but it has incorporated PV systems to simulate hybrid systems such as those proposed here, however it cannot optimise them. The NREL developed the program HOMER [19], which optimizes hybrid systems. This program uses the kinetic battery model [13]. The user must enter the parameters for the optimisation by choosing the different combinations for PV array power, the battery power and the inverter power. HOMER does not give the number of panels and their type as a solution, only a PV array power, from ones chosen by the user. The user must select the type of battery, and no optimization between different types of battery is made. Barley [14] has set a guideline about main dispatch strategies. Ohsawa [15] applied an artificial neural network to the operation control of PV-Diesel systems. Ashari [9] proposed the

optimisation of the dispatch strategy, based on Barley [14], by means of the Diesel generator stopping and starting set points. Kaiser [11] and El-Hefnawi [17] presented a method to design PV-Diesel systems. The optimization procedure starts by the definition of the model of the Diesel generator, and then optimizing the PV and battery

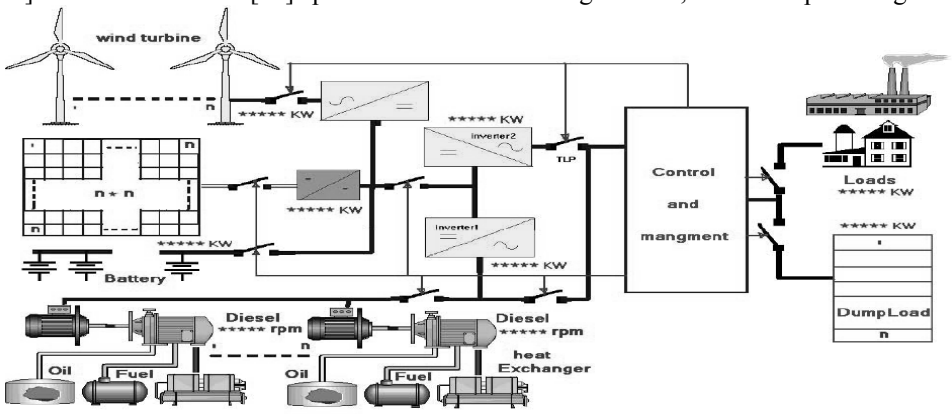



Fig. 1 Actual components of the proposed hybrid system

sizes, determining the minimum number of storage days and the minimum PV array area. The algorithm program described in this article, based on medium-penetration concept, improve hybrid wind - PV - diesel system using genetic algorithms.

## 2. Background

### 2.1 Renewable Penetration

When incorporating renewable-based technologies into remote stand-alone systems, the amount of energy that will be obtained from the renewable sources must be determined in order to properly size the added components. And because this will dictate which components will be used. Steve Drouilhet [18] developed the following classification and definitions of system penetration that characterize the levels of system complexity. A few criteria must be evaluated to determine the optimal hybrid configuration to size the wind turbine for this application. The percentage of renewable energy or renewable penetration can be classified in the following ways as per Drouihet [18]:

Instantaneous Penetration = ( Wind Power Output (kW) / Primary Electrical Load) (kW) *(1)*

Average Penetration = (Wind Power Energy Output (kWh) / Primary Electrical Load) (kWh) *(2)*

Instantaneous penetration is the ratio of how much power is being produced by the renewable resources at any specific instant and falls in the realm of the engineer. The average penetration is in the domain of the economist and includes a time domain thus it measured over days, months, or even years. A three level classification hybrid system based on system penetration that separates systems along power and system control needs.

### 2.2 Renewable energy sources models

To achieve an optimum reliability versus cost ratio in a hybrid system designs the share percentage of renewable energy sources in terms of system capacity is 70%-85% of load. In practice the share of renewable sources in a system would mostly be around 40%-60% [11].

#### 2.2.1 Available Wind Power (Pw)

The power of an air mass that flows at speed v, through a rotor disk of area A computed as in Eq. (3) ([20], [21]). Eq. (3) gives power in the wind, the actual power that can extract from the wind is significantly less than this figure suggests.

$$p_w = 0.5\rho A d V^3 \quad (3)$$

The theoretical optimum for utilizing the power in the wind by reducing its velocity was first discovered by Betz, in 1926. According to Betz, the theoretical maximum power, $P_{Betz}$, which can be extracted from the wind, is as shown Eq. (4).

$$P_{Betz} = 0.5\rho A d V^3 C_{pBetz} \quad (4)$$

Where: ρ = air density (kg/m3); v = wind speed (m/s); The power in the wind is proportional to the air density ρ, the intercepting area A, and the cubic of velocity v. Cp is coefficient of performance and has a value of 0.59. Thus, even if power extraction without any losses were possible, only 59% of the wind power could be utilized by a wind turbine.

#### 2.2.2 Solar energy

Solar energy is energy produced by the solar radiation, directly or in a diffuse way through the atmosphere. Because of various processes, it can be transformed into another form of useful energy for the human activity, in

particular in electricity or heat ([20],[21]). The maximum power provided by a solar panel is given by Eq. (5).

$$P_s = P_1.E_C[1+P_2(T_j - T_{jref})] \quad (5)$$

Where Ec is solar radiation, $T_{jref}$ is the reference temperature of the panels of 25°C, $T_j$ is the cells junction temperature (°C), P1 represent the characteristic dispersion of the panels and the value for one panel is included enters 0.095 to 0.105 and the parameter P2= -0.47%/C° ; is the drift in panels temperature. The addition of one parameter , P3 to the characteristic as shown in Eq. (6), gives more satisfactory results:

$$P_s = P_1.E_C[1+P_2(T_j - T_{jref})].(P_3+E_C) \quad (6)$$

This simplified model makes it possible to determine the maximum power provided by a group of panels for solar radiation and panel temperature given, with only three constant parameters P1, P2 and P3 and simple equation to apply. A thermal solar power station consists of a production of solar system of heat which feeds from the turbines in a thermal cycle of electricity production [30].

### 2.2.3 Diesel generator sizing and operation

The diesel should be sized to be able to meet full load. The renewable energy sources could then cover maintenance intervals or fuel shortage intervals. For systems around 250kW or larger, it might make sense to operate multiple diesels of different sizes. The resolution takes account of the fuel costs and reducing of the emissions of the polluting gases. The aim of real power economic dispatch (ED) is to make the generator's fuel consumption or the operating cost of the whole system minimal by determining the power output of each generating unit under the constraint condition of the system load demands [22]. The fundamental of the economic dispatch problem is the set of input - output characteristic of a power generating unit. The output of the generating unit will be designed by $P_G$, the megawatt net power output of the unit. In addition to the fuel consumption cost, the operating cost of a unit includes labour cost, maintenance cost, and fuel transportation cost. It is difficult to express these costs directly as a function of the output of the unit, so these costs are included as a fixed portion of the operating cost. The Characteristic of the generating unit is nonlinear. It is a convex curve, which is shown in Figure (2). It can be observed from the input - output characteristic of the generating unit that the power output is limited by the minimal and maximal capacity of the generating unit Eq. (7), that is, the maximal power output of the generating unit is determined by the design capacity or rate capacity of turbine or generator. Generally, the input - output characteristic of the generating unit is a quadratic function, as shown in Eq. (8).

$$PG\ min \le PG \le PG\ max \quad (7)$$

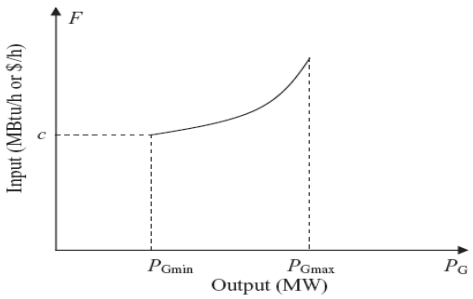



Fig.2 Input - output characteristic of the generating unit

$$F = a\,P^2_G + b\,P_G + c \quad (8)$$

Where a, b, and c are the coefficients of the input - output characteristic. The constant c is equivalent to the fuel consumption of the generating unit operation without power output, which is shown in above Figure (4).

## 3. Economics of Hybrid Systems Definition and Constraints

The use of hybrid off-grid electricity depends on the comparative costs, affordability, quality of service, and accessibility of other energy options which are locally available. This paper will concentrate on hybrid system design in terms of minimizing life cycle costs and also optimising the dispatch strategy while meeting a given demand reliably. Life cycle costs (LCC) are the sum of the equipment costs; the initial costs incurred at the beginning of a hybrid system electrification project; and discounted operation costs ;include running costs, maintenance and replacement costs; arising during the project until the end of the project horizon, which is usually set between 20 and 30 years. As shown in Figure (3). In The optimization problem the aim is to determine the new generation plants in terms of when to be available, what type and capacity

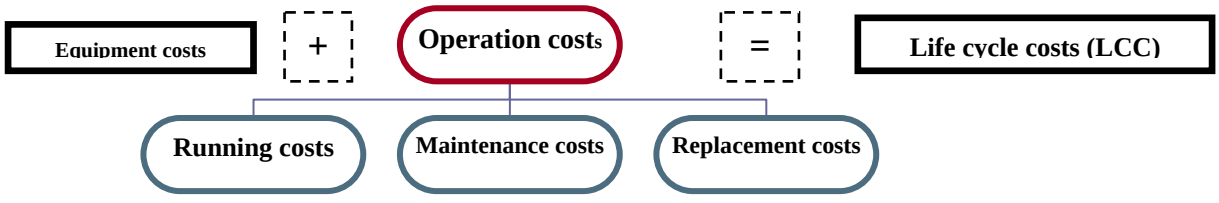



Fig.3 Life cycle costs

they should be and where to allocate so that an objective function is optimized and various constraints are met. It may be of static type in which the solution is found only for a specified stage (typically, year) or a dynamic type, in which, the solution is found for several stages in a specified period. The objective function or Life cycle costs (LCC) according to figure (3) consists of two term described in Eq.(9)

$$\text{Objective function} = \text{Capital costs} + \text{Operation costs} \quad (9)$$

The first term is, mainly due to Investment costs ($C_{inv}$), Salvation value of investment costs ($C_{salv}$) and Fuel inventory costs ($C_{finv}$) While the second term consists, mainly, of Fuel costs ($C_{fuel}$), Non-fuel operation and maintenance costs ($C_{O\&M}$) and Cost of energy not served ($C_{ENS}$). The objective function terms and the constraints are described in the following subsections.

### 3.1 Objective Functions

Total cost, $C_{total}$, to be minimized may be described in Eq.(10)

$$C_{total} = C_{inv} + C_{fuel} + C_{O\&M} + C_{ENS} \quad (10)$$

Where: $C_{inv}$ The investment cost, $C_{fuel}$ the fuel cost, $C_{O\&M}$ The operation and maintenance cost, $C_{ENS}$ The cost of energy not served and the details are as follows.

### 3.2 The Investment Cost

$X_{it}$ represents the number of unit type i required in year t, $C_{inv}$ is given by Eq.(11)

$$C_{inv} = \sum_{i=1}^{T} \sum_{i=1}^{NG} \text{Cost\_Inv}_{it} PG_i X_{it} \quad (11)$$

Where: $\text{Cost\_Inv}_{it}$ (The cost in \$/MW for unit type i in year t), $PG_i$( The capacity of unit i (MW)), T (The study period (in years)), NG(The number of units types)

### 3.3 The Fuel Cost

The fuel cost of each unit is a function of its energy output, normally in a nonlinear form. However, for simplicity, here we assume a linear function given by Eq.(12). Where

$$C_{fuel} = \sum_{i=1}^{T} \sum_{i=1}^{NG} \text{Cost\_Fuel}_{it} \text{Energy}_{it} X_{it} + \text{Cost\_Fuel}_{et} \quad (12)$$

$\text{Cost\_Fuel}_{it}$ The cost of fuel in \$/MWh for unit type i in year t), $\text{Energy}_{it}$(Energy output for unit type i in year t), $\text{Cost\_Fuel}_{et}$ (The fuel cost of existing units in year t)

### 3.4 The Operation and Maintenance Cost

Similar to $C_{inv}$, the operation and maintenance cost is given as a linear function of $PG_i$ given by Eq.(13)

$$C_{O\&M} = \sum_{i=1}^{T} \sum_{i=1}^{NG} \text{Cost\_O \& M}_{it} PG_i X_{it} \quad (13)$$

Where: $\text{Cost\_O\&M}_{it}$(The operation and maintenance cost (in \$/MW) for unit type i in year t)

### 3.5 The Cost of Energy not served

A generation unit may be tripped out in a rate given by its Forced Outage Rate (FOR). The so called Energy Not Served (ENS) cannot be made zero, but should be minimized as a cost term. It is given by Eq.(14)

$$C_{ENS} = \sum_{i=1}^{T} \text{Cost\_ENS}_t \text{ENS}_t \quad (14)$$

Where: $\text{Cost\_ENS}_t$(The cost of the energy not served in year t (\$/MWh)), $\text{ENS}_t$(The energy not served in year t (MWh)). Some constraints have to be observed during the optimization process such generation capacity which should be sufficient in meeting the load, Fuel Constraint and Fuel Pollution Constraint. Besides the objective function, some constraints should also be met. A simple constraint is the one which describes the available generating capacity to be greater than the load.

### 3.6 Fuel Constraint and Pollution cost function

The fuel cost function $C(P_G)$ in \$/h is represented by a quadratic function Eq. (15) ([26],[25]). The coefficients $a_i$,

$$C(P_G) = \sum_{i=1}^{NG} a_i P_{G_i}^2 + b_i P_{G_i} + c_i \quad (15)$$

$b_i$ and $c_i$ are appropriate to every production unit, $P_{Gi}$ is the real power output of i-th generator and NG is the number of thermal generators. The atmospheric emission can be represented by a function that links emissions with the power generated by every unit. The emission of $SO_2$ depends on fuel consumption and has the same form as the fuel cost [27]. The emission function in ton/h which represents $SO_2$ and $NO_x$ emission is a function of generator output and is expressed as follow in Eq.(16) [29].

$$E(P_G) = \sum_{i=1}^{NG} \alpha_i + \gamma_i P_{G_i}^2 + \beta_i P_{G_i} + \xi_i \exp\left(\lambda_i P_{G_i}\right) \quad (16)$$

Where $\alpha_i$, $\beta_i$ and $\gamma_i$ are the coefficients of emission function corresponding to the $i^{th}$ generator. These three parameters are determined by adjustment techniques of curves based on reel tests [28].

### 3.7 Problem constraints

The Production capacity constraints the generated real power of each generator at the bus i is restricted by lower limit max $P_{Gi}$ and upper limit min $P_{Gi}$ Eq. (17).

$$P_{Gi\,\min} \le P_{Gi} \le P_{Gi\,\max} \rightarrow i = 1,\ldots, N_G \quad (17)$$

The Power balance constraint the total power generation and the wind power must cover the total demand ' P' D and the power loss p in transmission lines Eq. (18).

$$P_D^{'} + P - \sum_{i=1}^{NG} P_{Gi} = 0 \qquad (18)$$

The Active power loss constraint the transmission and transport lines are positives Eq. (19). Renewable power

$$p > 0 \qquad (19)$$

constraint the renewable power used for dispatch should not exceed the 10 % of total power Demand Eq. (20),

$$0.10 \leq P_S + P_W \leq 0.85 P_D \qquad (20)$$

thus, the problem to be solved is formulated as follow Minimize of C ( $P_G$ ), E ( $P_G$ )

## 4. The Proposed Strategy and Control Algorithm

This paper mainly deals with hybrid system design and operation control problem which is non-linear due to non-linear component characteristics and the complexity of the hybrid system component interaction. The proposed system configuration is composed from six 250 kW wind turbine , PV arrays resulting in a total rated power of 200 kWp, and diesel generators sets arranged as follow six 250 kVA, four 100 kVA and ten 20 kVA acting as a backup system. When the PV and wind resources are not sufficient to supply the load demand (2000 kw) with its minimum operational level the diesel generators will recompense this demand. Based on the costs of components, fuel, labour, transport and

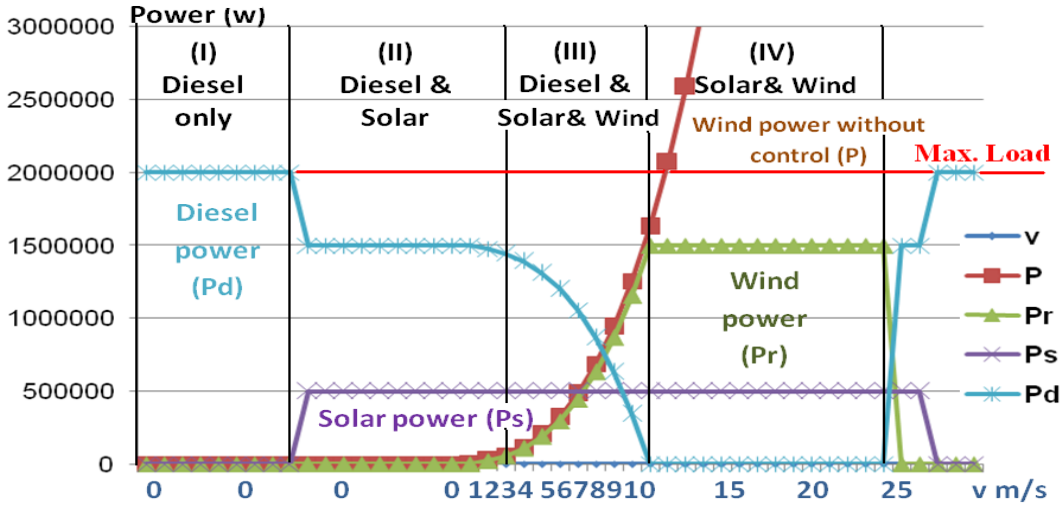



Fig.4. Hybrid system operation strategy

maintenance, the most cost-effective dimensions of all components and their operation strategy is evaluated as a target for the design algorithm. Operating the components effectively influences operation costs and, therefore, overall life cycle costs. The necessary optimization of the operation strategy in a hybrid system will focus on efficiency of diesel and prolonging component lifetimes. The proposed strategy for hybrid system operation is depicted as shown in Figure (4). It divided into four regions where region (I) represent the diesel unite operation only, region (II) represent the diesel and solar operation, region (III) represent the diesel, solar and wind turbine operation, finally region (IV) represent the operation of solar and wind turbine operation. Region (III) represents the management of demand and adjustment the renewable energy sources to maximizing the power and region (IV) represents the power regulation. The maximization of load factors is very important and has a significant influence on life cycle costs and sizing. The Control Algorithm Architecture, shown in Figure (5), for proposed hybrid system is developed according to the operation strategy conditions described in Figure (4). The renewable generators will reduce fuel consumption and engine generator maintenance. This is the task of the design optimization to recommend a least-cost and reliable design suitable for a given application with the aim to improve the system performance and lower costs. To develop the optimization problem of hybrid system design the electricity demand profile for a selected location with estimated weather conditions, costs for components, labour, transport and maintenance should be formulated. The demand is fixed to be 2 Mw in every design process and stage. Renewable energy resources, depend on the data of the climate such as the wind speed for wind energy, solar radiation and the temperature for solar energy, are calculated. So diesel power also calculated for each month of the year for each hour of the day as shown in Table 1.

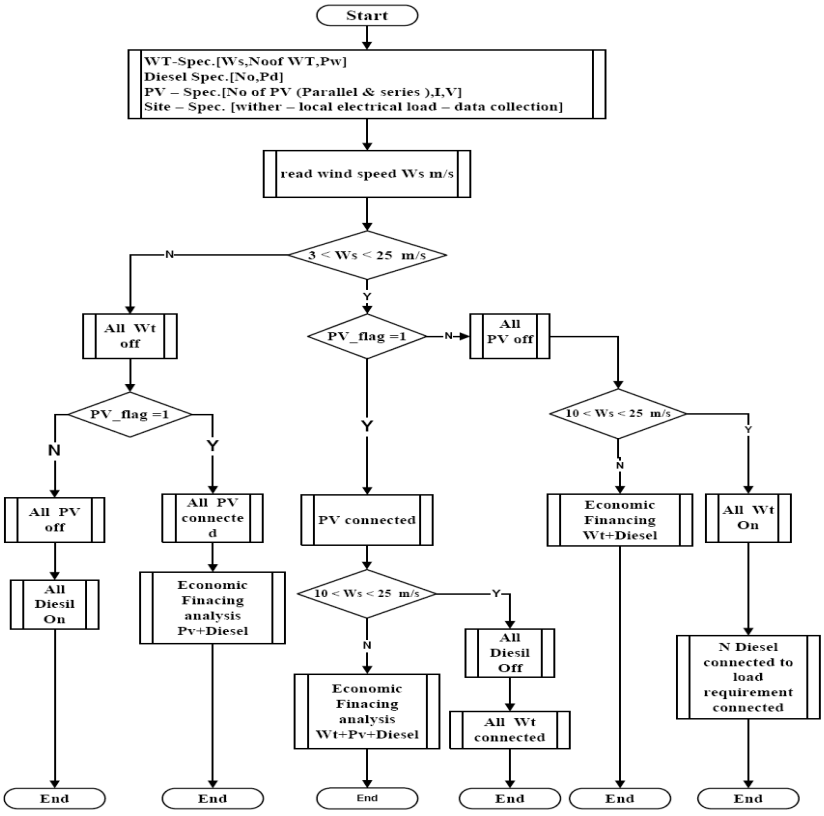



Fig.5. Control Algorithm Architecture for Hybrid System

Table 1: Average Diesel Power for each month of the year; calculated for each hour of the day

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1760000 | 1625000 | 1400000 | 1205000 | 950000 | 875000 | 935000 | 905000 | 845000 | 1115000 | 1535000 | 1700000 |
| 1 | 1790000 | 1655000 | 1460000 | 1265000 | 1010000 | 890000 | 965000 | 965000 | 905000 | 1205000 | 1595000 | 1715000 |
| 2 | 1790000 | 1685000 | 1520000 | 1310000 | 1055000 | 950000 | 1040000 | 1040000 | 950000 | 1250000 | 1580000 | 1715000 |
| 3 | 1790000 | 1700000 | 1535000 | 1385000 | 1100000 | 1040000 | 1115000 | 1100000 | 1010000 | 1310000 | 1595000 | 1715000 |
| 4 | 1790000 | 1715000 | 1595000 | 1460000 | 1115000 | 1055000 | 1160000 | 1115000 | 1100000 | 1355000 | 1595000 | 1700000 |
| 5 | 1790000 | 1700000 | 1595000 | 1505000 | 1160000 | 1115000 | 1220000 | 1205000 | 1145000 | 1385000 | 1625000 | 1700000 |
| 6 | 1290000 | 1230000 | 1125000 | 1020000 | 690000 | 570000 | 720000 | 690000 | 660000 | 915000 | 1125000 | 1215000 |
| 7 | 1275000 | 1215000 | 1170000 | 960000 | 615000 | 465000 | 615000 | 600000 | 600000 | 885000 | 1110000 | 1230000 |
| 8 | 1245000 | 1170000 | 1110000 | 885000 | 510000 | 390000 | 555000 | 495000 | 465000 | 750000 | 1020000 | 1230000 |
| 9 | 1200000 | 1065000 | 975000 | 855000 | 540000 | 390000 | 540000 | 450000 | 405000 | 660000 | 945000 | 1155000 |
| 10 | 1110000 | 1020000 | 915000 | 810000 | 540000 | 405000 | 600000 | 465000 | 405000 | 615000 | 900000 | 1110000 |
| 11 | 1065000 | 975000 | 855000 | 810000 | 510000 | 435000 | 615000 | 510000 | 405000 | 615000 | 855000 | 1035000 |
| 12 | 1020000 | 945000 | 840000 | 810000 | 495000 | 405000 | 600000 | 540000 | 405000 | 645000 | 840000 | 1005000 |
| 13 | 975000 | 900000 | 840000 | 795000 | 495000 | 405000 | 615000 | 510000 | 390000 | 645000 | 855000 | 945000 |
| 14 | 1005000 | 885000 | 840000 | 750000 | 495000 | 375000 | 645000 | 540000 | 390000 | 660000 | 885000 | 960000 |
| 15 | 1020000 | 900000 | 885000 | 810000 | 510000 | 375000 | 645000 | 540000 | 435000 | 705000 | 945000 | 975000 |
| 16 | 1080000 | 1005000 | 945000 | 885000 | 555000 | 405000 | 660000 | 570000 | 495000 | 795000 | 1020000 | 1065000 |
| 17 | 1170000 | 1095000 | 1020000 | 915000 | 540000 | 450000 | 690000 | 615000 | 555000 | 915000 | 1110000 | 1185000 |
| 18 | 1260000 | 1155000 | 1080000 | 975000 | 510000 | 465000 | 705000 | 660000 | 570000 | 885000 | 1125000 | 1200000 |
| 19 | 1730000 | 1655000 | 1535000 | 1415000 | 935000 | 890000 | 1145000 | 1055000 | 950000 | 1205000 | 1580000 | 1685000 |
| 20 | 1715000 | 1580000 | 1415000 | 1310000 | 785000 | 785000 | 905000 | 830000 | 815000 | 1055000 | 1505000 | 1655000 |
| 21 | 1700000 | 1565000 | 1355000 | 1160000 | 725000 | 725000 | 800000 | 755000 | 755000 | 1040000 | 1505000 | 1655000 |
| 22 | 1730000 | 1520000 | 1340000 | 1145000 | 785000 | 725000 | 800000 | 755000 | 755000 | 1055000 | 1475000 | 1655000 |
| 23 | 1745000 | 1565000 | 1355000 | 1160000 | 845000 | 800000 | 845000 | 815000 | 785000 | 1115000 | 1535000 | 1685000 |
| PdC (Kw)= | 34045 | 31525 | 28705 | 25600 | 17470 | 15385 | 19135 | 17725 | 16195 | 22780 | 29860 | 32890 |

Pdc (Power of Diesel Consumed) Kw = Load demand Kw – (wind power + solar power ) Kw

## 5. The Proposed Economic Genetic Algorithms

Genetic algorithms (GA) are used in this paper to solve the economic dispatch problem and to optimize the developed hybrid system design model through minimizing its life cycle costs while still meeting required system performance. The developed Genetic algorithms Architecture for proposed economic dispatch is shown in Figure (6). GA provides a solution to a problem by working with a population of individuals each representing a possible solution. Each possible solution is termed a "chromosome." New points of the search space are generated through GA operations, known as reproduction, crossover, and mutation. These operations consistently produce fitter offspring through successive generations, which rapidly lead the search toward global optima.

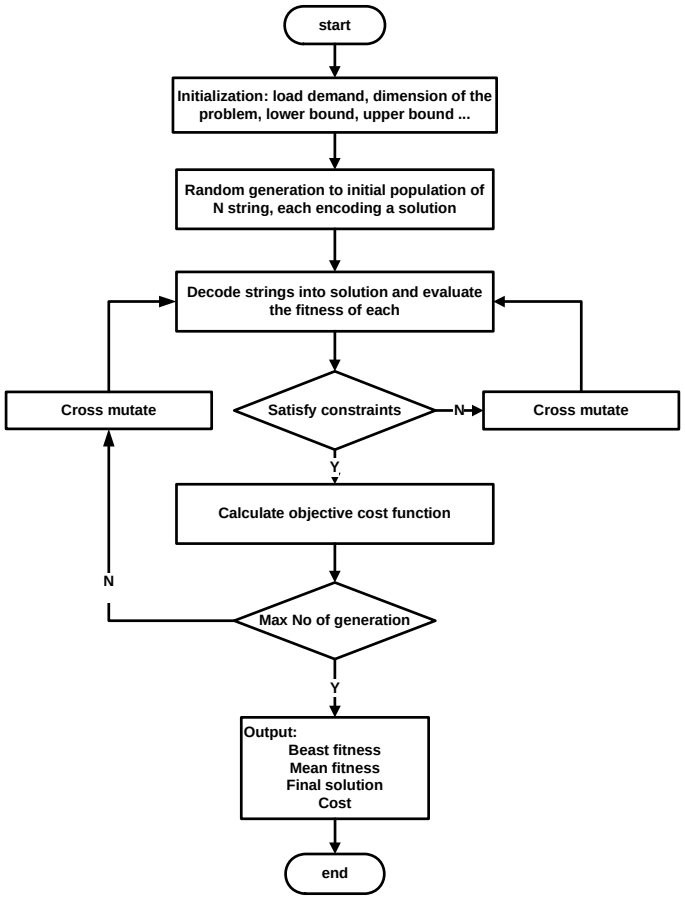



Fig.6. Genetic algorithms for economic dispatch

The economic dispatch (ED) problem can be stated as below in Eq. (21) and Eq. (22). The outputs of the (N – 1)

$$\min \quad F = \sum_{i=1}^{N} Fi\ (P_{Gi}) \qquad (21)$$

$$\sum_{i=1}^{N} P_{Gi} = P_D \qquad (22)$$

free generators can be chosen arbitrarily within limits while the output of the reference generator is constrained by the power balance. It is assumed that the N $^{th}$ generator is the reference generator. These (N − 1) strings are concatenated to form a consolidated solution bit string of 8 * (N − 1) bits called the genotype. A population of m genotypes must be initially generated at random. Each genotype is decoded to a power output vector. The output of the reference unit is described in Eq. (23). Adding penalty factors h1, h2 to the violation of power output of

$$P_{GN} = P_D - \sum_{i=1}^{N} P_{Gi} \qquad (23)$$

the slack bus unit; we can combine the above equations in Eq. (24).

$$F_a = \sum_{i=1}^{N} Fi(P_{Gi}) + h_1(P_{GN} - P_{GN\max})^2 + h_2(P_{GN\min} - P_{GN})^2 \qquad (24)$$

Where $P_{GN\ min}$, $P_{GN\ max}$ are the lower and upper limits of the power output of the Slack bus unit, respectively. The value of the penalty factors should be large so that there is no violation for unit output at the final solution. The
GA fitness function is defined as the inverse of ($F_a$) Eq. (24) as in (25).

$$F_{fitness} = \frac{1}{F_A} \qquad (25)$$

The unit step size can be computed by Eq. (26). Where n is the length of the substring in binary codes corresponding to a unit.

$$S_i = \frac{P_{Gi\ \max} - P_{Gi\ \min}}{2^n - 1} \qquad (26)$$

## 6. Results

The optimization using genetic algorithms has the following parameters Total load = 350000, Dimension of problem or No. of diesels (6*250kw,4*100kw,5*20kw) = 15 units, Max No of iterations = 1000, 25 .
Population size = 100 and give the following results

### 6.1 Final solution

D1 = 24.99154, D2= 24.98670, D3 = 24.98876, D4 = 24.08463, D5 = 24.98483, D6 = 24.99022, D7 = 24.28500, D8 = 23.08660, D9 = 24.05695, D10 = 33.54844, D11= 19.99843, D12 = 19.99944, D13 = 19.99942,
D14 = 19.99955, D15 = 19.99949e+004] kw.
Associated cost: = 7.267849e+007

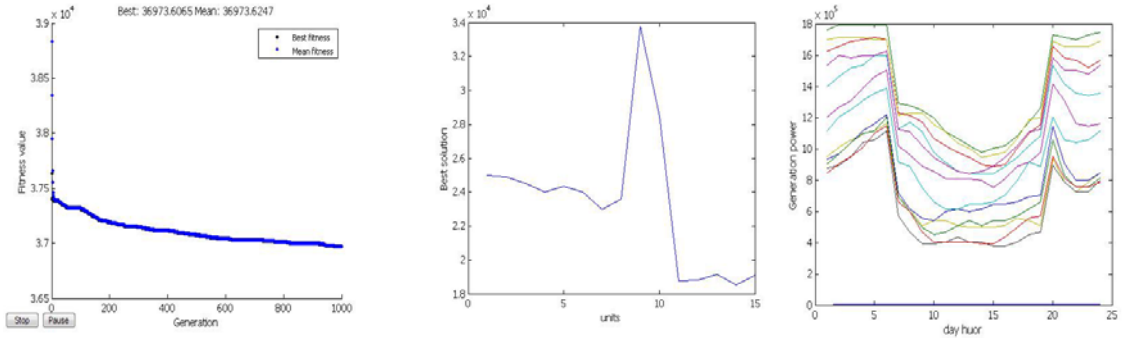



Fig.6. Control Algorithm results from lift to right. (a) Best fitness, (b) best solution, (c) generation power

Decreasing Max No of iterations from 1000 to 20 make the algorithm more faster than previous and give the following results with Total load = 450000 and Max No of iterations = 20

6.2 Final solution is

[ D1 = 34.74849, D2= 34.74840, D3 = 34.88047, D4 = 27.4924, D5 = 33.84436, D6 = 34.75015, D7 = 34.77807, D8 = 34.70787, D9 = 34.74953,
D10 = 46.69673, D11= 19.74678, D12 = 19.74588, D13 = 19.68406, D14 = 19.70812, D15 = 19.72366] kw

Associated cost: = 1.267935e+008

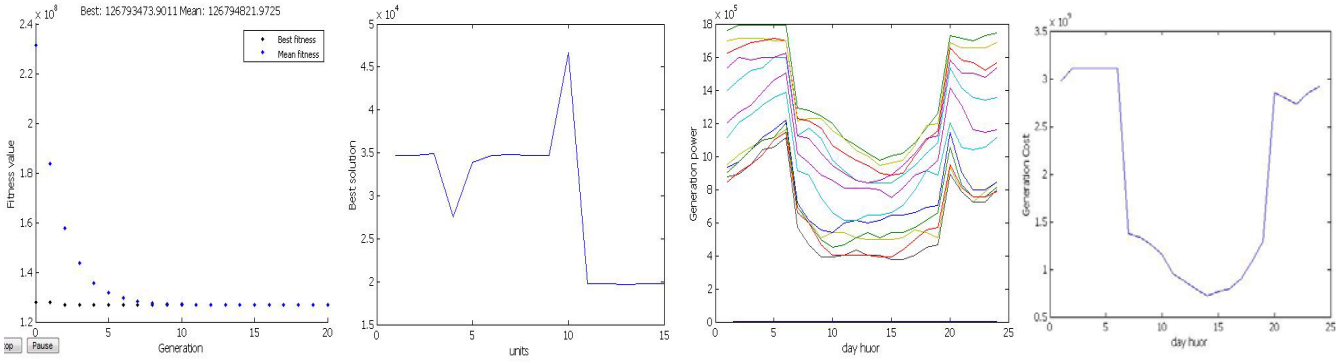



Fig.7. results from lift to right. (a) Best fitness, (b) best solution, (c) generation power,(d) generation cost for one day

6.3 Finally the Associated cost for monthly average hours diesel power generation: cost =1.0e+009 *

Columns 1 through 6

2.7357 2.7961 2.7961 2.7961 2.7357 2.7357

Columns 7 through 12

1.1876 1.2237 1.2237 1.0506 0.9557 0.8192

Columns 13 through 18

0.7690 0.6724 0.6959 0.7199 0.8722 1.1176

Columns 19 through 24

1.1522 2.6761 2.5591 2.5591 2.5591 2.6761

## 7. Conclusion

The algorithm described in this paper, based on medium-penetration concept, improve hybrid (wind - PV - diesel) system using genetic algorithms. The program calculates the optimum configuration of the system according to the weather data as well as the period of operation of solar cells, either morning or evening. Hybrid system is a combination of diesel power available continuously and are available locally, and pollution-free wind power, solar energy which is one of the advantages of this system. Where the annual diesel fuel consumption reduced and the pollution minimized at the same time. The control algorithm takes full advantage of the wind energy and solar energy when it is available and minimizes diesel fuel consumption. A hybrid energy (wind-PV-Diesel) system has greater reliability for electricity production than a PV-only system or wind only (Diesel engine production is independent of atmospheric conditions). This provides greater flexibility, higher efficiency and lower costs for the same energy quantity produced. Also, PV-Diesel systems, compared with Diesel-only systems, provide a reduction of the operation costs and air pollutants emitted to the atmosphere.

## REFERENCES

[1] Miranda, V. and Hang, P. S. (2005), *Economic dispatch model with fuzzy constraints and attitudes of dispatchers*,

IEEE Transactions on Power Systems, Vol. 20, No. 4, Nov., pp.2143-2145, 2005.

[2] Lingfeng Wang and Chanan Singh., 2006, *Tradeoff between Risk and Cost in Economic Dispatch Including Wind Power Penetration Using Praticle Swarm Optimisation*, International Conference on Power System Technology, 2006.

[3] Zhao et al./ JZhejiang Univ SCI 2005. , *Multiple objective particle swarm optimization technique for economic load dispatch*, 6A(5) :420-427.

[4] Abido M. A. and Bakhashwain J. M., 2005, *Optimal VAR dispatch using a multiobjective evolutionary algorithm. Electrical power and energy systems*, PP. 13-20, 2005.

[5] Abido M. A., 2003, *A niched Pareto genetic algorithm for multi objectives environmental/economic dispatch*, Electrical power and energy systems, PP. 97-105, 2003

[6] Wang, L. F. and Singh, *C. Multi-objective stochastic power dispatch through a modified particle swarm optimization algorithm*, Special Session on Applications of Swarm Intelligence to Power Systems, Proceedings of IEEE Swarm Intelligence Symposium,Indianapolis, May, pp. 127-135, 2006.

[7] T. BOUKTIR, L. SLIMANI and M. BELKACEMI *,A Genetic Algorithm for Solving the Optimal Power Flow Problem*, Leonardo Journal of Sciences, Issue 4, January-June p.44-58, 2004.

[8] Benjamin Baran, Member, IEEE, José Vallejos, Rodrigo Ramos and Ubaldo Fernandez, Member, IEEE. ,*Reactive Power Compensation using a Multi-objective Evolutionary Algorithm*, PPt 2001, IEEE Porto Power Tech Conference, 10th-13th September, Porto,Portugal.

[9] Ashari, M., Nayar, C.V., 1999, *An Optimum Dispatch Strategy Using set Points for a Photovoltaic (PV)-Diesel-Battery Hybrid Power System*, Solar Energy, Vol. 66, No 1, pp.1-9.19

[10] Wichert, B., 1997, *PV-Diesel hybrid energy systems for remote area power generation a review of current practice and future developments*, Renewable and Sustainable Energy Reviews, Vol. 1, No. 3, September 1997, pp. 209-228.

[11] Seeling-Hochmuth, G.C., 1998, *Optimisation of hybrid energy systems sizing and operation control*, A Dissertation presented to the University of Kassel in Candidacy for the Degree of Dr.-Ing.

[12] Green, H.J., Manwell, J., 1995, *HYBRID2 – A Versatile Model of the Performance of Hybrid Power Systems*, Proceedings of WindPower'95, Washington DC, March 27-30, 1995.

[13] Manwell, J.F., McGowan, J.G., 1993, *Lead acid battery storage model for hybrid energy systems*, Solar Energy, Vol. 50, pp.399-405.

[14] Barley, C.D., Winn, C.B., Flowers, L., Green, H.J., 1995, *Optimal Control of Remote Hybrid Power Systems*, Part I: Simplified Model. Proceedings of WindPower'95, Washington DC, March 27-30, 1995.

[15] Ohsawa, Y., Emurd, S. and Arai, K., 1993, *Optimal operation of photovoltaic / Diesel power generation system by neural network*, Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, 1993. ANNPS '93., 19-22 April 1993. pp. 99-103

[16] Kaiser, R, Sauer, D.U., Armbruster, A., Bopp, G., Puls, *H.G. New Concepts for System Design and Operation Control of Photovoltaic Systems*, Proc. 14th European Photovoltaic Solar Energy Conference, Barcelona, June 1997

[17] El-Hefnawi, Said H., 1998, *Photovoltaic Diesel-generator hybrid power system sizing*, Renewable Energy, Vol. 13, No. 1, pp. 33-40.

[18] Drouihet, S., *High-Penetration AC Bus Wind-Diesel Hybrid Power Systems*, NREL, Village Power '98 Technical Workshop

[19] Piwko, R., Osbom, D., Gramlich, R., Jordan, G., Hawkins, D., and Porter, K. (2005), *Wind energy delivery issues*, IEEE Power & Energy Magazine, November/December, pp. 67-56, 2005.

[20] Lingfeng Wang, Chanan Singh, 2007, *Compromise Between Cost and Reliability in Optimum Design of An Autonomous Hybrid Power System Using Mixed-Integer PSO Algorithm,* Department of Electrical and Computer Engineering Texas A&M University

[21] Faisal A. Mohamed, Heikki N. Koivo., (2007), *Online Management of MicroGrid with Battery Storage Using Multiobjective Optimization*, POWERENG 2007, April 12-14, Setubal, Portugal, 2007

[22] J. Nanda , R.B. Narayanan , *Application of genetic algorithm to economic load dispatch with Line - flow constraints ,* Electrical Power and Energy Systems , Vol. 24 ,2002 , pp. 723 – 729 .

[23] G.B. Sheble , K. Brittig , *Refined genetic algorithm - economic dispatch example* , IEEE Trans on Power System , Vol. 10 , No. 1 , 1995 .

[24] Farag A., Al-Baiyat S. and Cheng TC., 1995, *Economic load dispatch multi objectives optimization procedures using linear programming techniques*, IEEE Trans. On Power Syst., Vol. 10, No.2, PP. 731-738, 1995.

[25] Abido M. A., 2003, A niched Pareto, genetic *algorithm for multi objectives environmental/economic dispatch.* Electrical power and energy systems, PP. 97-105, 2003

[26] D. B. Das and C. Patvardhan, *New Multi-objective Stochostic Serch Technique for Economic Load Dispatch*, 'IEE Proc.-Gener. Transm. Distrib., Vol. 145,No. 6, pp.747- 752, 1998.

[27] Talaq J., El-Hawary F. and El-Hawary M., 1994, *A summary of Environmental/Economic Dispatch Algorithms. Trans. On Power Systems*, Vol. 6, No. 3, Aug 1994, pp. 1508-1516, 1994.

[28] DeMeo, E. A., Grant, W., Milligan, M. R., and Schuerger, M. J. (2005), *Wind plant integration: costs, status, and issues,* IEEE Power & Energy Magazine, November/December, pp. 38-46, 2004.

[29] Zahavi J., Eisenberg L., *Economic–environmental power dispatch. IEEE Trans. Syst*., Vol. 5, No. 5, 1985, PP. 485–489, 1985.

[30] Saoussen B., Hsan H. ABDALLAH, and Abderrazak O., *Economic Dispatch for Power System included Wind and Solar Thermal energy*, Leonardo Journal of Sciences, ISSN 1583-0233, p. 204-220, 2009

**Osama Abdel Hakeem Abdel Sattar** is a Production Research Center Manager in A.O.I.E.F. He received his B.Sc., from Faculty of Engineering, Department of Electrical Engineering, (Electronics and Communication section) Helwan University, Egypt, his Dipl.-Ing. degree (in Control Eng.) and M.Sc.( in computer engineering) in 1993, 2001 and 2005,respectively. He became a Teacher in the Ministry Of Education in 1997, and R&D Engineer in A.O.I in 2000, and VLSI Designer in A.O.I in 2005, and wind turbine Designer in A.O.I in 2008 until now. He is a member of the Society and HF Technology, DigChip Member ,DriverGuide Member.

**R. R. Darwish** received the B.Sc. and M.Sc degrees in Electronics and Communications Engineering from Helwan University, Egypt in 2000 and 2004, respectively. She has obtained her Ph.D in the field of wireless sensor network from the Department of Electronics and Communications Engineering at Helwan University in 2009. Currently she is an associative professor in the Department of Mechatronics at the Faculty of Engineering, Helwan University, Egypt. Her research interests include image processing, wireless sensor networks, and cloud computing.

**Saad Mohamed Ali Eid** is a Professor of Control Engineering, Faculty of Engineering, Univ. of Cairo. He received his B.Sc. degree and M.sc. degree in Electronics Engineering (Communication section) from Cairo Univ., his Dr.-Ing degree from Stuttgart Univ. , West Germany ,at 1963,1968 and 1973 respectively . He became an Associate Prof. and a Professor in 1978, and 1983 respectively. He was an International scientific member of the ECCTD, 1983. He is Author of 70 and/or Co-author of 200 scientific papers.

**Elsayed Mostafa Saad** is a Professor of Electronic Circuits, Faculty of Engineering, Univ. of Helwan. He received his B.Sc. degree in Electrical Engineering (Communication section) from Cairo Univ., his Dipl.-Ing. degree and Dr.-Ing degree from Stuttgart Univ. , West Germany ,at 1967,1977 and 1981 respectively . He became an Associate Prof. and a Professor in 1985, and 1990 respectively. He was an International scientific member of the ECCTD, 1983. He is Author and/or Co-author of 132 scientific papers. He is a member of the national Radio Science Comittee, member of the scientific consultant committee in the Egyptian Eng. Syndicate for Electrical Engineers, till l May 1995, Member of the Egyptian Eng. Sydicate, Member of the European Circuit Society (ECS), Member of the Society of Electrical Engineering(SEE).

# Quantitative Multiscale Analysis using Different Wavelets in 1D Voice Signal and 2D Image

**Niraj Shakhakarmi**

**Department of Electrical & Computer Engineering, Prairie View A&M University (Texas A&M University System)**
**Prairie View, Houston, Texas, 77446, USA**

**Abstract**

Mutiscale analysis represents multiresolution scrutiny of a signal to improve its signal quality. Multiresolution analysis of 1D voice signal and 2D image is conducted using DCT, FFT and different wavelets such as Haar, Deubachies, Morlet, Cauchy, Shannon, Biorthogonal, Symmlet and Coiflet deploying the cascaded filter banks based decomposition and reconstruction. The outstanding quantitative analysis of the specified wavelets is done to investigate the signal quality, mean square error, entropy and peak-to-peak SNR at multiscale stage-4 for both 1D voice signal and 2D image. In addition, the 2D image compression performance is significantly found 93.00% in DB-4, 93.68% in bior-4.4, 93.18% in Sym-4 and 92.20% in Coif-2 during the multiscale analysis.

***Keywords:*** *Quantitative, Multiscale Analysis, Different Wavelets, One Dimensional Voice Signal, Two Dimensional Image.*

## 1. Introduction

First generation wavelets transform essentially needs the Fourier transform and the basis functions which are dyadically scalable with translation property of one particular mother basis function. These are the first non-trivial wavelets developed around 1980s. These include the Daubechies wavelet, Haar wavelet, Shannon Wavelet, Coiflets Wavelet and the Meyer wavelet. The major drawback of the first generation wavelet is that it can be deployed for infinite or periodic signals and cannot be optimized in the bounded domain. These wavelets transforms (WTs) are used in identifying pure frequencies, de-noising signals, detecting discontinuities and breakdown points, detecting self-similarity and compressing images.

Second generation wavelets transform originates with concept of Lifting scheme to maintain the time-frequency localization and fast algorithms instead of fourier domain to deploy in geometrical applications. This should replace translation and dilation as well as any Fourier analysis. The basic algorithm of the lifting scheme, is to split up even samples then are adjusted to serve the coarse version of the original signal data in even set and odd set dilation as well as any Fourier analysis. The basic algorithm of the lifting scheme is to split even samples then are adjusted to serve the coarse version of the original signal data in even set and odd set then predict odd signal using even part to detect the missing parts called details and update even samples for adjustment to serve the coarse version of the original signal. These WTs are extensively used for lossy data compression, in geographical data analysis, computer graphics and efficient coding in compression algorithm.

Third generation wavelets transform are the complex wavelet transform (CWT) with the complex-valued extension to the standard discrete wavelet transform (DWT). It is typically two-dimensional wavelet transform deployed for the multi-resolution, sparse representation, and useful feature characterization based on the structure of an image. The major pros are that these WTs do not exhibit oscillations, lack of directivity, aliasing and degree of shift-variance in its magnitude. But, the major cons are that it exhibits two dimension of the signal being transformed and yields the redundancy compared to a separable.

Next generation wavelets transform optimize the PSNR, error free, lossless and advanced multi level resolution. These wavelets will be more advanced in terms of efficiency and performance. These WTs are still under research and they will focus specific applications such as human vision characterization, frequency localization, feature extraction, seismic analysis, bio-medical analysis and so on.

Multiscale analysis represents the hierarchy of structural implementation to enhance the physical characteristics of the signal (both 1D and 2D). When the multiscale stage (level) is increased then it provides the fine resolution from coarse resolution. In other words, it is the systematic process to analyze signal at lower multiscale stage with coarse resolution and then higher multiscale stage with fine resolution [1-2]. Thus, higher stage of the multiscale using wavelets provides significant multiresolution improving signal quality. This is deployed using different

kinds of wavelets in signal decomposition and reconstruction to investigate their performance at stage-3 and stage-4. The Haar wavelet, Deubechies wavelet, Morlet wavelet, Cauchy wavelet, Shannon wavelet, DCT, FFT, Biorthogonal wavelet, Symmlet wavelet and Coiflet wavelet are deployed in 1D signal and 2D image in this paper.

## 2. Problem & Proposed Solution

The problem is to analyze and compare the quantitative mutiscale features of different wavelets transform and determine the qualitatively suitable wavelets on 1D voice signal and 2D image for multi-resolution. This is addressed by decomposition and reconstruction of 1D voice signal and 2D image by deploying different wavelets transform at the third and fourth multi-resolution stage. In addition, the quantitative analysis of 1D signal and 2D image is done in terms of SNR, MSE, entropy and PSNR for different wavelets transform.

The proposed solution includes the following aspects:

- 1D Signal decomposition and reconstruction at stage-4 using different wavelets
- Quantitative analysis of 1D Signal at stage-4
- 2D Image decomposition and reconstruction at stage-4 using different wavelets
- Quantitative analysis of 2D Image at stage-4
- 2D Image compression at stage-4

### 2.1 One Dimensional Signal Decomposition & Reconstruction

One dimensional discrete wavelet transform is used for 1 D signal decomposition and reconstruction in time-scale (frequency) representation of non-stationary signals. It is based on multi-resolution approximation in which a function uses scaling function at various resolutions so that the lost details can be recovered using wavelet functions and the original signal is reconstructed by adding approximation and detail coefficient. It is deployed by a sequence of low pass and high pass filters [3-6]. Low pass (LP) filters are associated with the scaling function and provide approximation whereas high pass (HP) filters are associated with the wavelet function and provide detail lost in approximating the signal.

#### 2.1.1 1D Signal Analysis (Decomposition) at Stage-4 using Different Wavelets

1D signal decomposition is done using a sequence of LP and HP filter banks at four different stages by cascading at LP downsample decimated by 2 as shown in Fig-1.

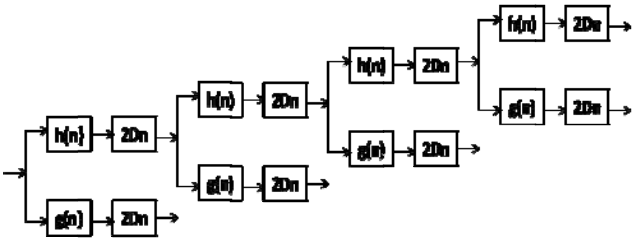



Fig. 1 1D Signal Decomposition at stage- 4

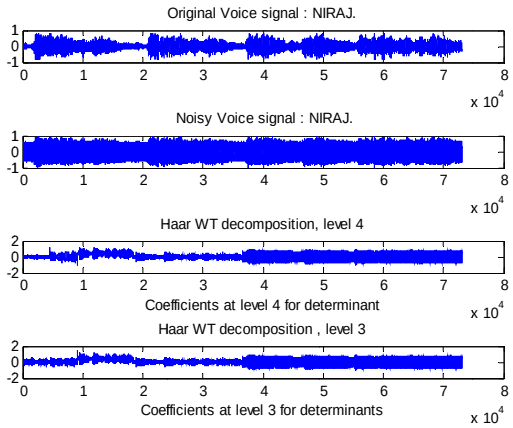



Fig. 2 Haar WT Decomposition at stage- 4

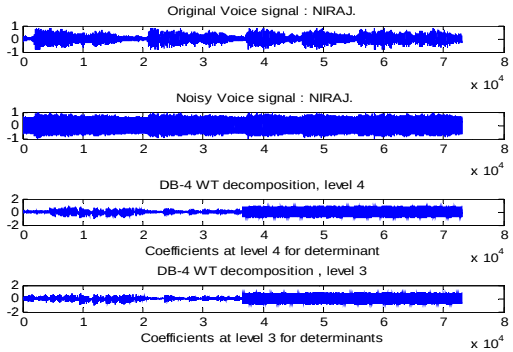



Fig. 3 DB-4 WT Decomposition at stage- 4

1D voice signal is decomposed at stage-3 & stage-4 using DCT, FFT and different wavelets such as Haar, Deubachies, Morlet, Cauchy, Shannon, Biorthogonal, Symmlet and Coiflet as illustrated in Fig. 2 to Fig. 12.

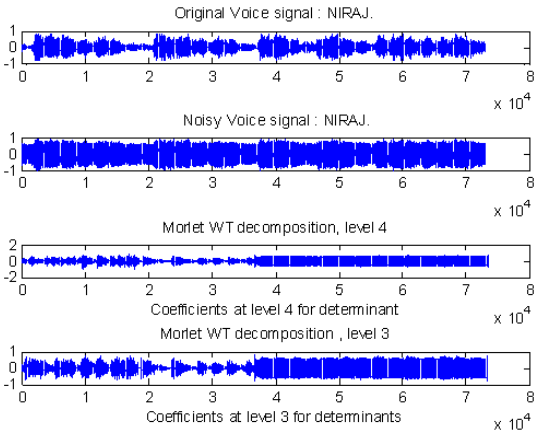



Fig. 4 Morlet WT Decomposition at stage- 4

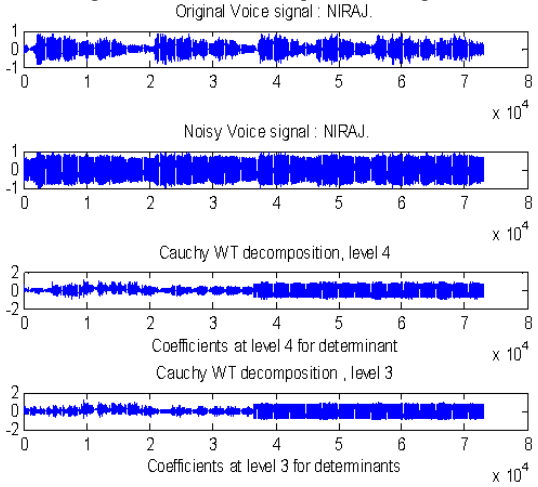



Fig. 5 Cauchy WT Decomposition at stage- 4

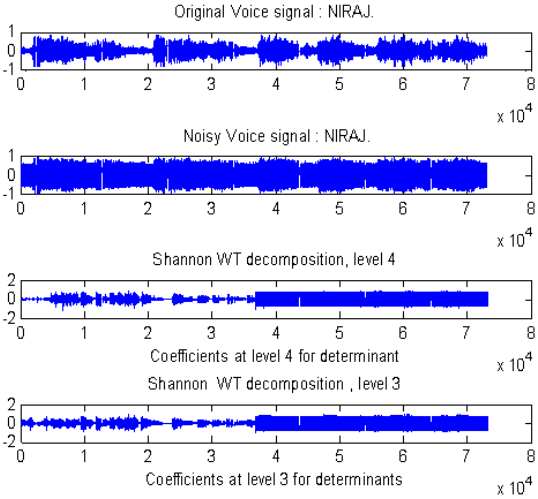



Fig. 6 Shannon WT Decomposition at stage- 4

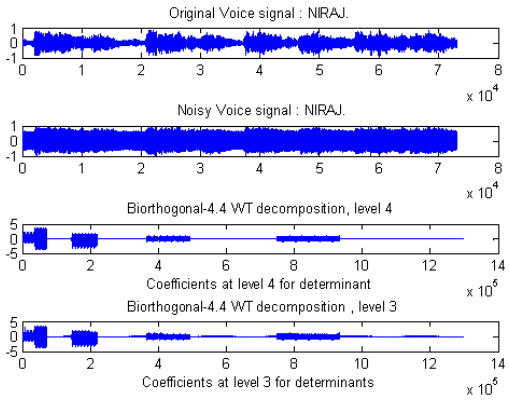



Fig. 7 Biorthogonal WT Decomposition at stage- 4

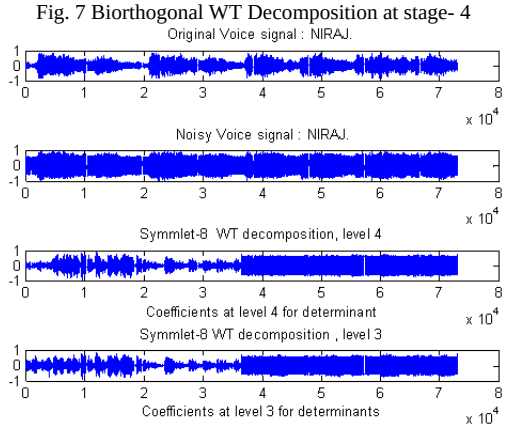



Fig. 8 Symmlet-8 WT Decomposition at stage- 4

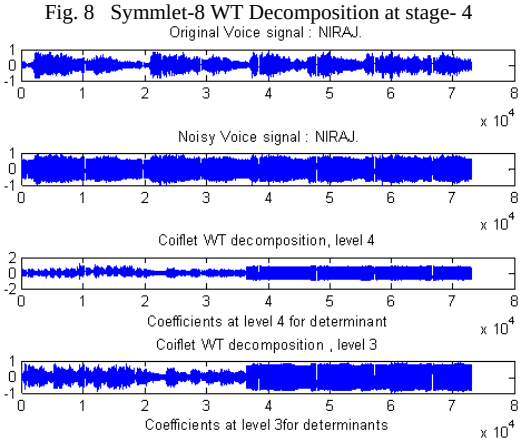



Fig. 9 Coiflet WT Decomposition at stage- 4

### 2.1.2 One Dimensional Signal Reconstruction (Synthesis) Stage-4 using Different Wavelets

1D signal reconstruction is done using a sequence of LP and HP filter banks at four different stages by cascading at LP upsample decimated by 2 as shown in Fig-10.

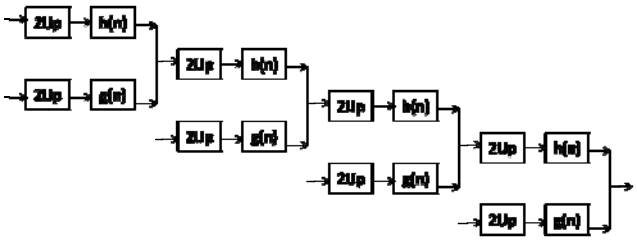



Fig. 10 1D signal Reconstruction at stage- 4

1D voice signal is reconstructed at stage-3 & stage-4 using DCT, FFT and different wavelets such as Haar, Deubachies, Morlet, Shannon, Biorthogonal, Symmlet and Coiflet as illustrated in Fig. 11 to Fig. 19.

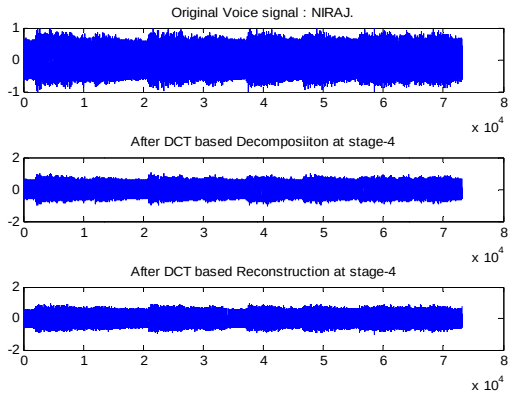



Fig. 11 DCT Analysis and Synthesis at stage-

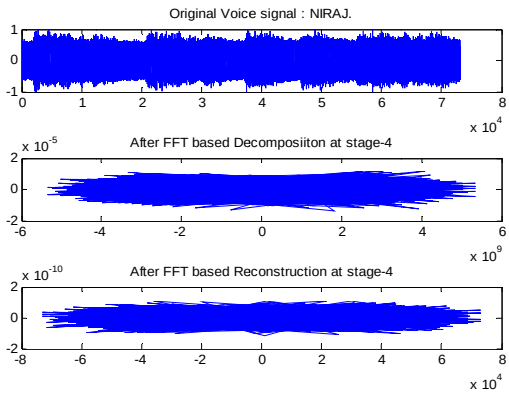



Fig. 12 FFT Analysis and Synthesis at stage- 4

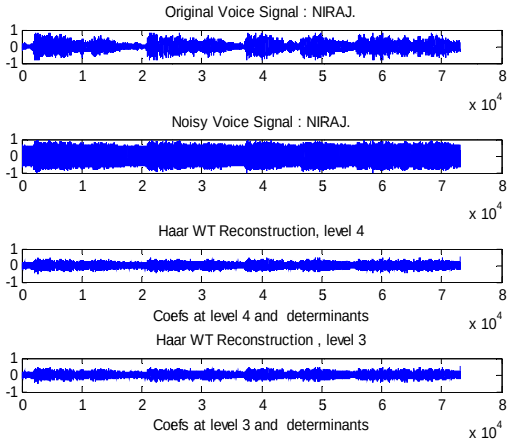



Fig. 13 Haar WT Reconstruction at stage- 4

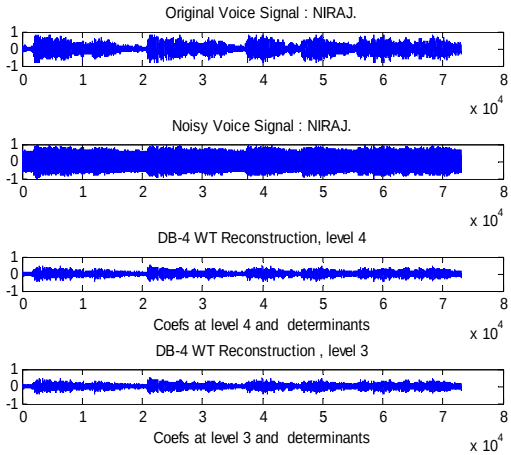



Fig. 14 DB-4 WT Reconstruction at stage- 4

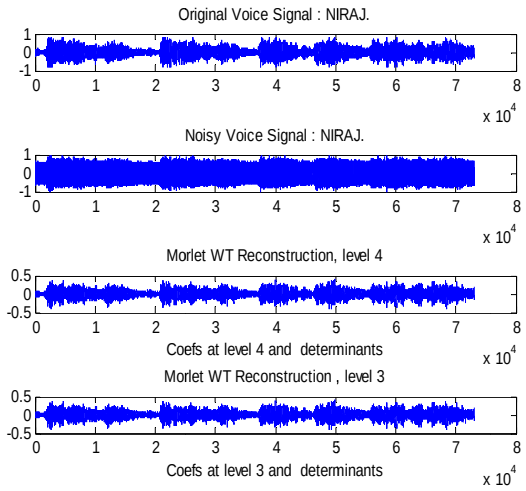



Fig. 15 Morlet WT Reconstruction at stage- 4

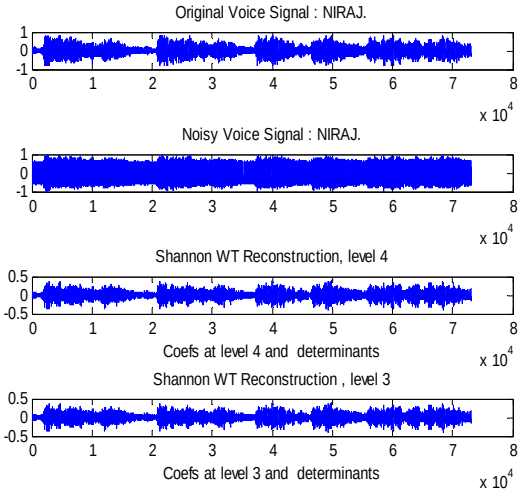



Fig. 16 Shannon WT Reconstruction at stage- 4

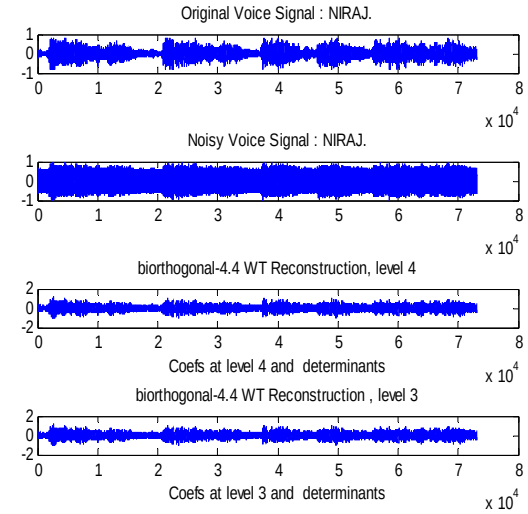



Fig. 17 Biorthogonal WT Reconstruction at stage-4

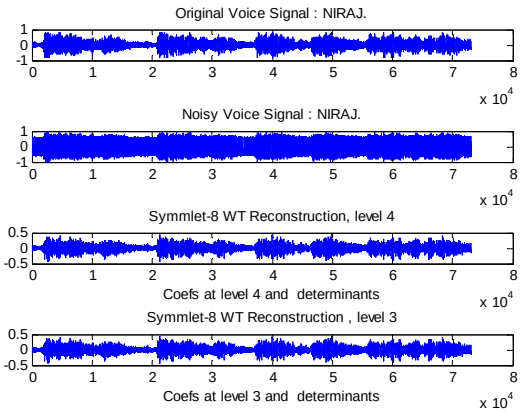



Fig. 18 Symmlet-8 WT Reconstruction at stage-4

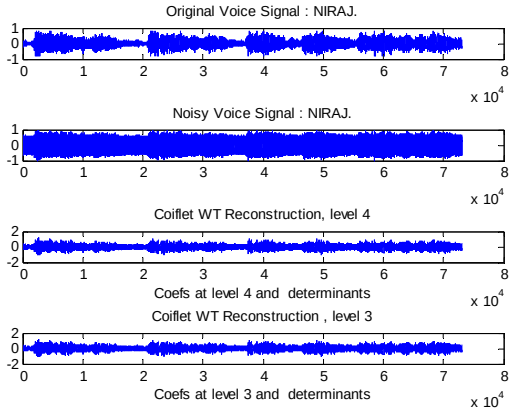



Fig. 19 Coiflet WT Reconstruction at stage- 4

### 2.1.3 Quantitative Analysis of Different Wavelets on 1D Noisy Voice Signal at stage-4

Table 1: Quantitative Analysis of Different Wavelets on Noisy Voice Signal at stage-4

| Different Wavelets | SNR (db) | MSE | Entropy | PSNR (db) |
|---|---|---|---|---|
| **Haar** | 81.39 | 1.4058e-034 | 4.1246 | 92.6011 |
| **DB-4** | 60.06 | 1.9053e-012 | 3.8457 | 28.7194 |
| **Morlet** | 71.65 | 1.3231e-013 | 3.7496 | 17.1357 |
| **Cauchy** | 52.01 | 1.2170e-011 | 3.9384 | 36.7727 |
| **Shannon** | 74.20 | 7.3524e-014 | 3.7494 | 14.5841 |
| **DCT** | 24.69 | 6.5711e-009 | 4.7578 | 64.0962 |
| **FFT** | 97.27 | 1.0349e+004 | 4.6210 | 86.07 |
| **Biorthogonal-2.4** | 72.67 | 1.9379e-006 | 4.4204 | 88.7932 |
| **Biorthogonal-4.4** | 75.35 | 1.9358e-006 | 4.4407 | 88.7885 |
| **Symmlet-8** | 67.54 | 3.4066e-013 | 3.7608 | 21.2431 |
| **Coiflet** | 52.01 | 1.2170e-011 | 3.9384 | 36.7727 |

From the quantitative analysis on voice signal at stage-4, it is found that Haar provides highest SNR, as well as lower MSE. Similarly, DCT and Bior-4.4 provide highest Entropy and Haar provides best PSNR. The average histogram is found approximately 7.3113e+003 in all cases. Specifically, Haar WT does not have overlapping windows, and reflects only changes between adjacent sample pairs. The Haar wavelet uses only two scaling and wavelet function coefficients, thus calculates pair wise averages and differences. That's why, Haar is found best

WT for noisy voice signal decomposition and reconstruction at stage-4.

## 2.2 Two Dimensional Image Decomposition and Reconstruction

2D discrete wavelet transform is used for 2D image decomposition and reconstruction using 2D scaling and wavelet functions which are orthogonal to its own translation [1-2], [5-6]. It consists of four sets of coefficients which are known as approximation coefficients, detail coefficients along the horizontal direction, detail coefficients along the vertical direction, detail coefficients along the diagonal direction.

### 2.2.1 2D Image Analysis (Decomposition) at stage-4 using Different Wavelets

2D image decomposition is done using a sequence of combination of LP and HP filter banks in rows and columns (LL, LH, HL, HH) at four different stages by cascading at LL downsample decimated by 2 as shown in Fig. 20 and Fig. 21.

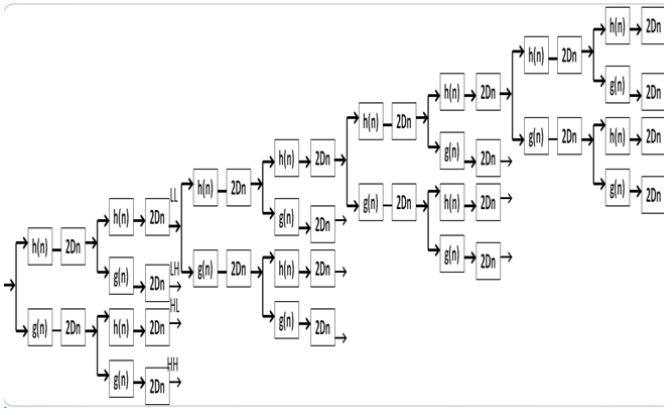



Fig. 20 2D Image Decomposition at stage- 4

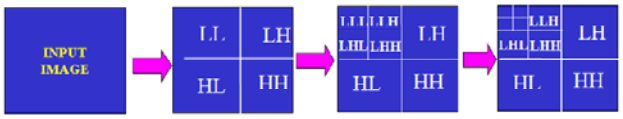



Fig. 21 Multiscale Analysis of 2D Image at stage-3

2D finger print image is decomposed at stage-2, stage-3 & stage-4 using DCT, FFT and different wavelets such as Haar, Deubachies, Morlet, Biorthogonal, Symmlet and Coiflet as illustrated in Fig.29, Fig.30 and Fig. 22 to Fig. 27.

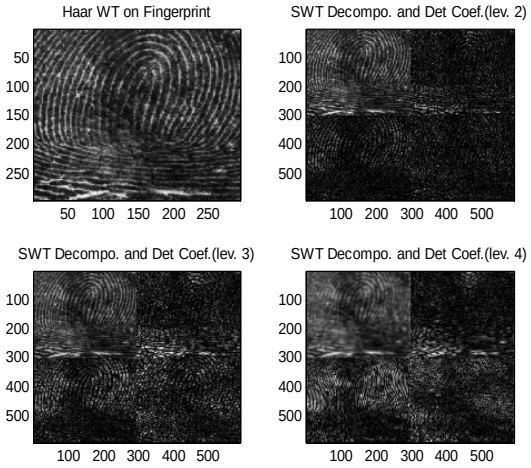



Fig. 22 Haar WT Decomposition at stage -4

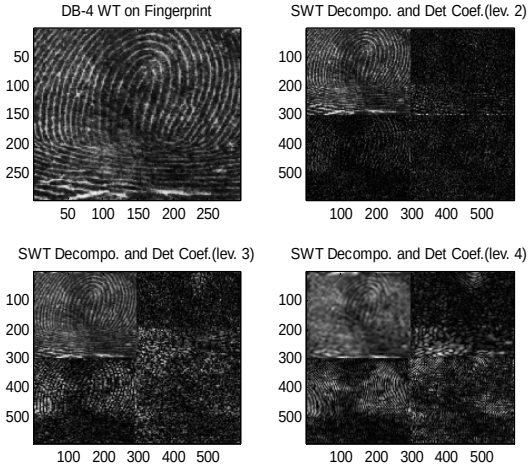



Fig. 23 DB-4 WT Decomposition at stage -4

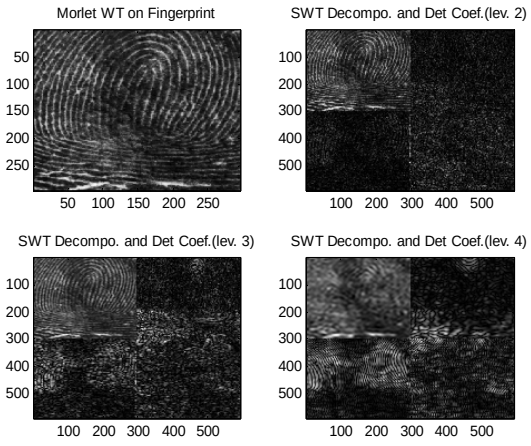



Fig. 24 Morlet WT Decomposition at stage -4

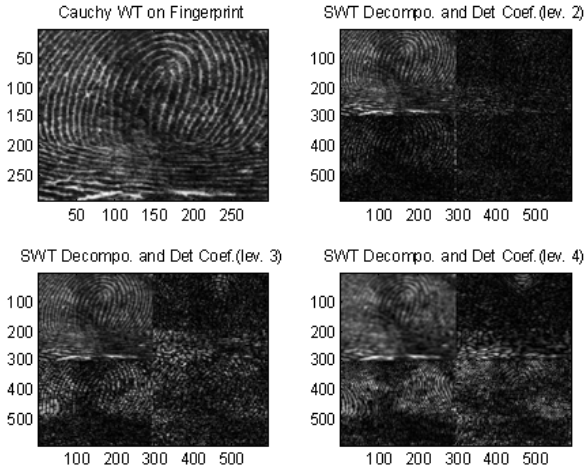



Fig. 25 Cauchy WT Decomposition at stage -4

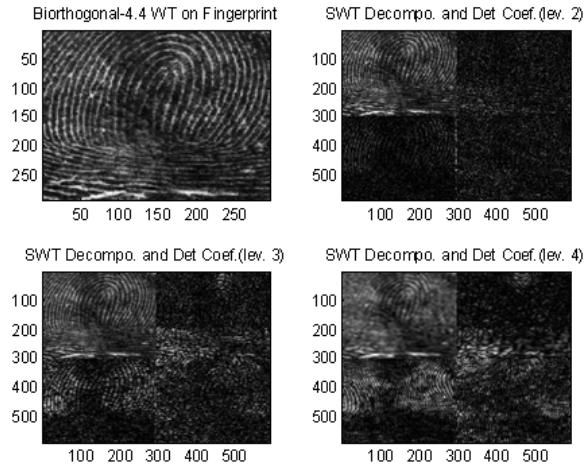



Fig. 26 Bior-4.4 WT Decomposition at stage -4

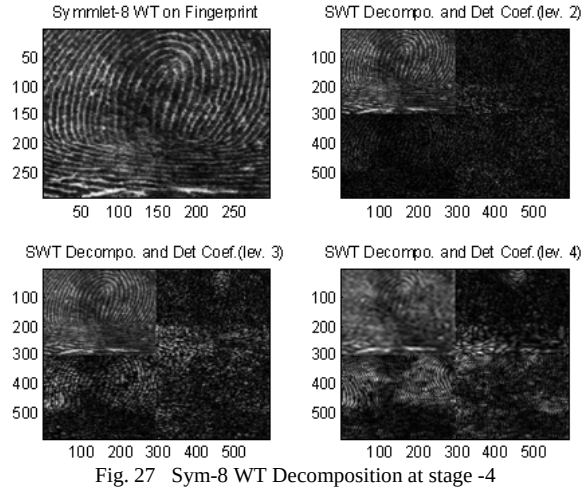



Fig. 27 Sym-8 WT Decomposition at stage -4

### 2.2.2 2D Image Synthesis (Reconstruction) at stage-4 using Different Wavelets Transform

2D image reconstruction is done using a sequence of combination of LP and HP filter banks in rows and columns (LL, LH, HL, HH) at four different stages by cascading at LL upsample decimated by 2 as shown in Fig. 28.

2D finger print image is reconstructed at stage-2, stage-3 & stage-4 using DCT, FFT and different wavelets such as Haar, Deubachies, Biorthogonal and Symmlet as illustrated in Fig. 29 to Fig. 35.

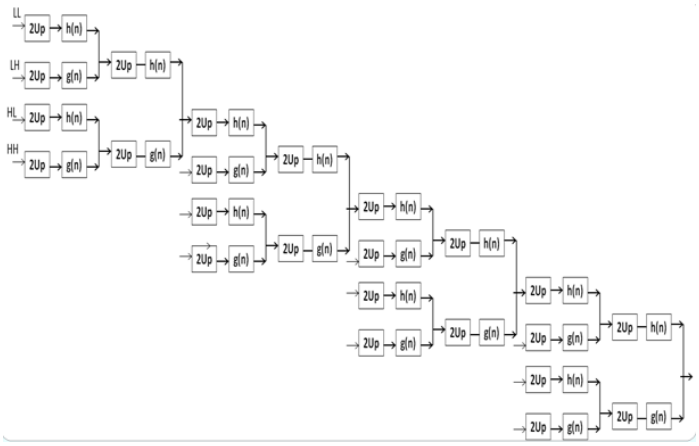



Fig. 28 2D Image Reconstruction at stage- 4

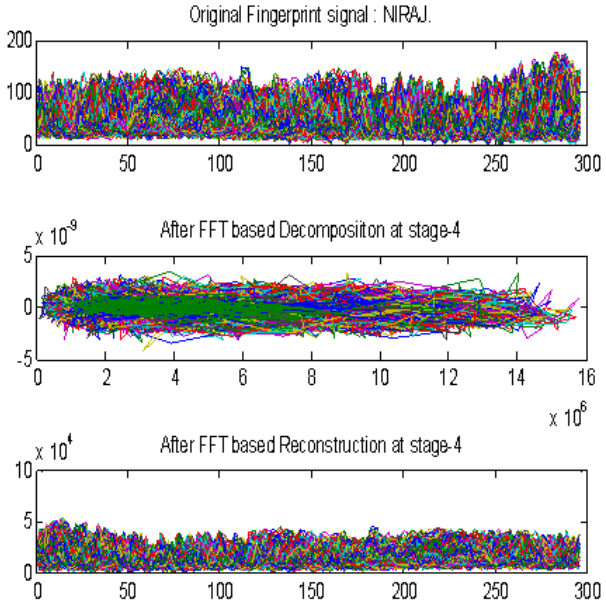



Fig. 29 FFT Analysis & Synthesis at stage -4

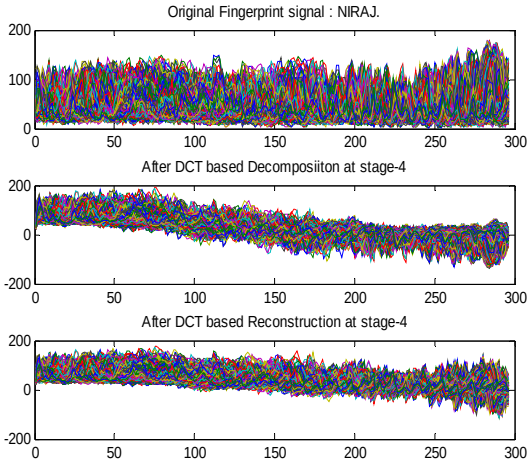



Fig. 30 DCT Analysis & Synthesis at stage -4

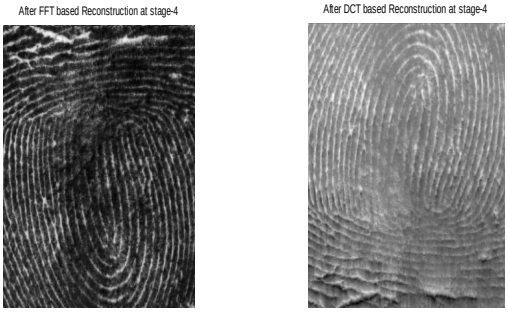



Fig. 31 Image Synthesis by FFT & DCT at stage -4

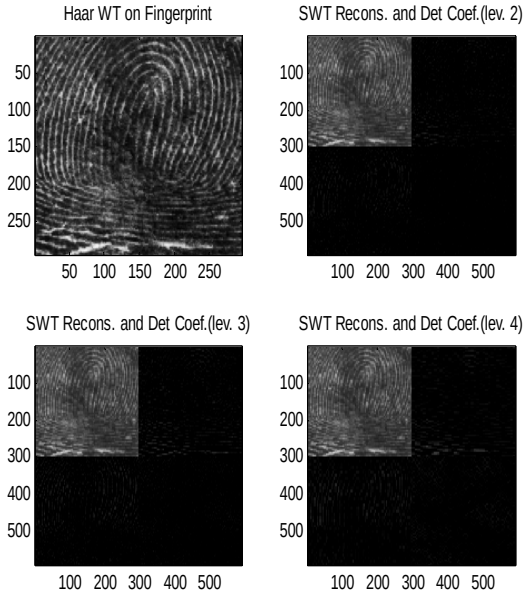



Fig. 32 Haar WT Reconstruction at stage -4

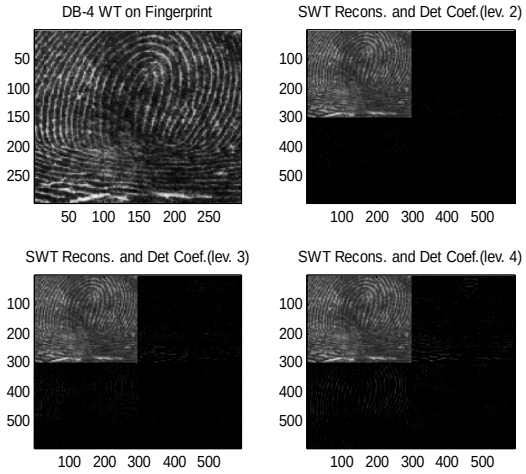



Fig. 33 DB-4 WT Reconstruction at stage -4

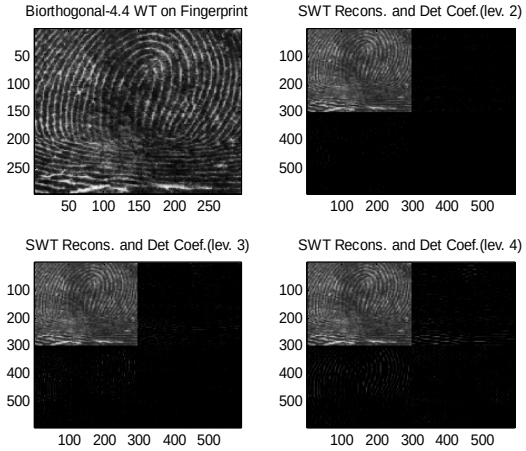



Fig. 34 Bior-4.4 WT Reconstruction at stage -4

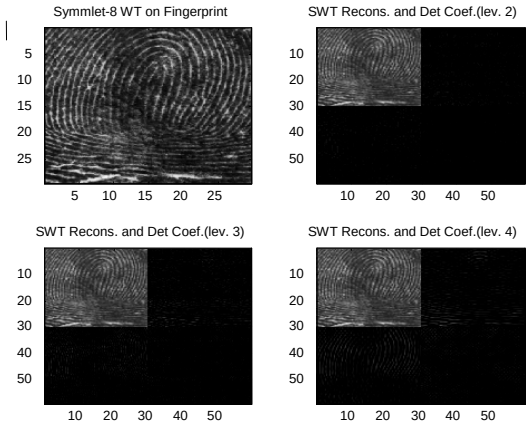



Fig. 35 Sym-8 WT Reconstruction at stage -4

### 2.2.3 Quantitative Analysis of Different Wavelets on 2D Image at stage-4

Table 2: Quantitative Analysis of Different Wavelets on 2D fingerprint image at stage-4

| Different Wavelets | SNR (db) | MSE | Entropy | PSNR (db) |
|---|---|---|---|---|
| **Haar** | 7.3167 | 8.4974 | 1.6483 | 28.03 |
| **DB-4** | 8.4560 | 7.4169 | 1.9496 | 26.89 |
| **Morlet** | 8.9697 | 6.6969 | 1.7505 | 26.38 |
| **DB-2** | 10.300 | 6.8423 | 2.0726 | 25.05 |
| **Cauchy** | 7.5217 | 6.3821 | 1.5782 | 21.03 |
| **Shannon** | 9.5147 | 6.6731 | 1.7338 | 25.83 |
| **DCT** | 7.9153 | 36.606 | 0.5591 | 26.13 |
| **FFT-2** | 8.139 | 37.576 | 0.5129 | 25.36 |
| **Biorthogonal-2.4** | 12.552 | 6.1138 | 1.9846 | 22.80 |
| **Biorthogonal-4.4** | 12.448 | 6.3087 | 1.8981 | 22.90 |
| **Symmlet-8** | 8.1908 | 7.3157 | 1.8272 | 27.16 |
| **Coiflet** | 10.300 | 6.8423 | 1.5726 | 25.05 |

From above quantitative analysis on 2D Fingerprint Image at stage-4, it is found that highest SNR and lowest MSE in Biorthogonal-2.4, higher MSE in Haar WT except FFT-2 & DCT. Similarly DB-2 as well as Bior-2.4 provides the highest Entropy and Haar as well as Sym-8 provides best PSNR. The average histogram is found 12.80 in all cases.

## 3. Two Dimensional Image Compression at stage-4

The salient compression performance is found 93.00% in DB-4, 93.68% in bior-4.4, 93.18% in Sym-4 and 92.20% in Coif-2 deploying hard threshold=30 at 4 stage analysis on the fingerprint image as illustrated in Tables 3-6.

Table 3: Compression performance using wavelets

| Daubechies | DB-4 | DB-6 | DB-8 | DB-10 |
|---|---|---|---|---|
| **Compression** | 93.00 | 92.21 | 91.04 | 90.08 |

Table 4: Compression performance using wavelets

| Biorthogonal | Bior2.4 | Bior4.4 |
|---|---|---|
| **Compression** | 91.47 | 93.68 |

Table 5: Compression performance using wavelets

| Symmlet | Sym4 | Sym8 | Sym10 | Sym25 |
|---|---|---|---|---|
| **Compression** | 93.18 | 91.53 | 90.49 | 84.49 |

Table 6: Compression performance using wavelets

| Different Wavelets | Coif2 | Coif5 | Rbio5.5 | Dmey |
|---|---|---|---|---|
| **Compression** | 92.20 | 88.34 | 89.70 | 76.59 |

Daubechies wavelets are seen compactly supported and have highest number of vanishing moments whereas Biorthogonal wavelets compactly supported wavelets for symmetry and exact reconstruction. The time taken is not linear for wavelet decomposition and reconstruction. On the other hand, Symmlets are compactly supported wavelets with highest number of vanishing moments and the best compression is given by Sym-8 [6-8].

The original image and compressed image deploying different wavelets such as Sym-8, Bior-2.4and Bior-4.4 and Db-4 are illustrated in Fig. 36 and Fig. 37.

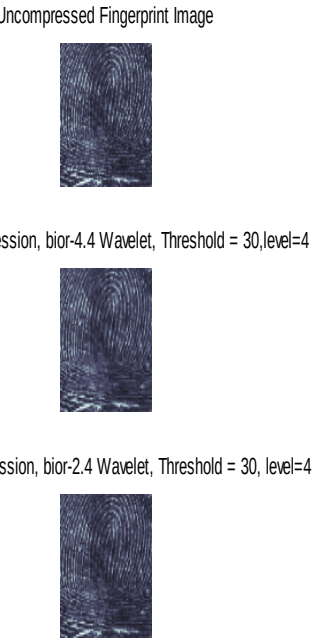



Fig. 36 Image compression by Bior- 4.4 & Bior- 2.4

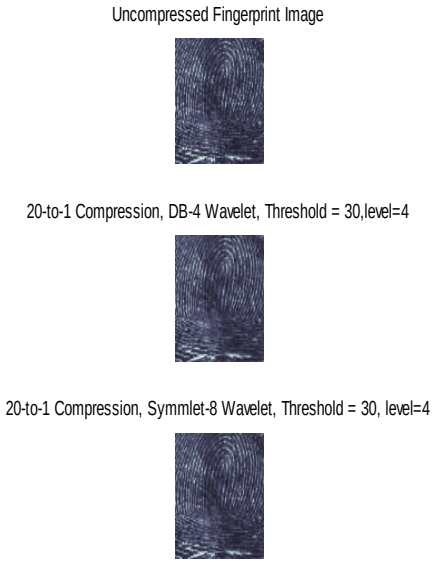



Fig. 37 Image compression using DB-4 and Sym-8

## 4. Conclusion

Multiscale analysis concludes that MSE is increased with the increasing number of stages whereas the SNR decreases. The quantitative analysis on 1D noisy voice signal at stage-4, shows that highest SNR and lower MSE in Haar, highest Entropy in DCT and Bior-4.4 and best PSNR in Haar. On the other hand, the quantitative analysis in 2D fingerprint image at stage-4, concurs that highest SNR and lowest MSE in Biorthogonal-2.4, higher MSE in Haar WT except FFT-2 & DCT. Similarly DB-2 as well as Bior-2.4 provides highest Entropy and Haar as well as Sym-8 provides best PSNR. Furthermore, the simulation results show the significant image compression. The salient compression performance is found that 93.00% in DB-4, 93.68% in bior-4.4, 93.18% in Sym-4 and 92.20% in Coif-2 deploying hard threshold=30 at stage 4 on the 2D fingerprint image. Future research will concentrate the application of next generation of wavelets on video frames.

## References

[1] S.G. Mallat, "A Theory for Multiresolution Signal Decom -position : The Wavelet Representation", IEEE Transaction On pattern Analysis and Machine Intelligence, Vol.2, No.7, July 1989.

[2] M.J. Shensa,"The discrete wavelet transform: Wedding the á trous and Mallat algorithms", IEEE Transaction on Signal Processing, Vol. 10, pp. 2463–2482, 1992.

[3] G. Beylkin, and N. Satio, "Wavelets, their Autocorrelation functions and multi resolution representation of signals", IEEE Trans. Signal Processing, Vol. 7, pp. 147–164, 1997.

[4] E. J. Cand`es, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", IEEE Transaction on Information Theory, Vol. 52, No. 2, pp. 489–509, 2006.

[5] W. L. Chan, H. Choi, and R.G. Baraniuk, "Coherent Multiscale Image Processing Using Dual-Tree Quaternion Wavelets", IEEE Transaction On Image processing , Vol.17, No.7, July 2008.

[6] W.S. Lin, S.K. Tjoa, H. J. Zhao, and K. J. Liu, "Digital Image source coder forensics via Intrinsic fingerprints", IEEE Transaction on Information Forensics and Security, Vol. 4, No. 3, pp. 460, Sept. 2009.

[7] C. M. Stamm, and L.K. Ray, "Wavelet based Image Compression antiforensics", In Proceedings of 2010 IEEE 17th International Conference on Image Processing Sep 26-29, 2010.

[8] P. L. Dragotti , and M. Vetterli, "Wavelets Footprints: theory, algorithms, and applications", IEEE Transaction On Signal Processing, Vol. 51, pp.1-18, May 2003.

**Dr. Niraj Shakhakarmi** worked as a Doctoral researcher

from 2009 to 2011 in the US ARO (Army Research Office) funded Center for Battlefield Communications (CeBCom) Research, Department of Electrical and Computer Engineering, Prairie View A&M University (Texas A&M University System). He received his B.E. degree in Computer Engineering in 2005 and M.Sc. in Information and Communications Engineering in 2007. He has accomplished Ph.D in Electrical & Computer Engineering in 2011 from Prairie View A&M University, Houston, USA. His research interests are in the areas of Wavelets applications and Digital Image Processing, Secured Position Location & Tracking (PL&T), Cognitive Radio Networks, Mobile Ad hoc Networks, 4G Networks, Satellite Networks, QoS & Network Security, and Color Technology. He has published WSEAS journal, ICSST conference, Elsevier conference, and several papers are under review in IEEE and WSEAS journals and conference papers.

IJCSI

ISSN (online): 1694-0814

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

## CALL FOR PAPERS International Journal of Computer Science Issues (IJCSI)
## Volume 9, Issue 3 – May 2012 Issue

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

**Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Cornell University Library, ScientificCommons, EBSCO, ProQuest and more.**

**Deadline: 31st March 2012**
**Notification: 30th April 2012**
**Revision: 10th May 2012**
**Publication: 31st May 2012**

**Context-aware systems**
**Networking technologies**
**Security in network, systems, and applications**
**Evolutionary computation**
**Industrial systems**
**Evolutionary computation**
**Autonomic and autonomous systems**
**Bio-technologies**
**Knowledge data systems**
**Mobile and distance education**
**Intelligent techniques, logics and systems**
**Knowledge processing**
**Information technologies**
**Internet and web technologies**
**Digital information processing**
**Cognitive science and knowledge**
**Agent-based systems**
**Mobility and multimedia systems**
**Systems performance**
**Networking and telecommunications**
**Software development and deployment**
**Knowledge virtualization**
**Systems and networks on the chip**
**Knowledge for global defense**
**Information Systems [IS]**
**IPv6 Today - Technology and deployment**
**Modeling**
**Software Engineering**
**Optimization**
**Complexity**
**Natural Language Processing**
**Speech Synthesis**
**Data Mining**

**For more topics, please see http://www.ijcsi.org/call-for-papers.php**

For more information, please visit the journal website (www.IJCSI.org)

www.IJCSI.org

# IJCSI

The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit ongoing research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

**Indexing of IJCSI**

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons
10. DBLP
11. EBSCO
12. ProQuest